Pictures in our heads: Visual stereotypes affect social categorization

Plaatjes in onze hoofden: Visuele stereotypen beïnvloeden sociale categorisatie

Ron Dotsch

ISBN: 978-90-8559-302-7

Cover illustration: Jeroen Derks

Printed by: Optima

Pictures in our heads: Visual stereotypes affect social categorization

een wetenschappelijke proeve op het gebied
van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen,
op gezag van de rector magnificus prof. mr. S. C. J. J. Kortmann,
volgens het besluit van het college van decanen
in het openbaar te verdedigen op Dinsdag 18 januari 2011
om 13:30 uur precies

door

Ron Dotsch
geboren op 11 juni 1980
te Amstelveen

Promotores:

Prof. dr. Daniël H. J. Wigboldus

Prof. dr. Ad van Knippenberg

Manuscript committee:

Prof. dr. Ap Dijksterhuis

Dr. Rob W. Holland

Prof. dr. Russell Spears (Cardiff University)

To opa Berry, oma Mary, saba Israel and safta Betty.

Preface

In 2008, I visited Opatija, Croatia, for a conference organized by the European Association of Social Psychology. It happened to coincide with the European Football Championship. On Monday, June 9th, the Dutch team would play against Italy. My Dutch colleagues in Opatija were extremely anxious for the match. So were our Croatian hosts. After exploring Opatija for some time, I headed back to our hotel together with Maarten Bos (Dutch social psychologist, tall, blonde, blue eyes, cleanly shaven), Rob Holland (Dutch social psychologist, tall, blonde, blue eyes, cleanly shaven), and Daniël Wigboldus (Dutch social psychologist, tall, blonde, blue eyes, cleanly shaven). I, on the other hand, am a Dutch social psychologist, not that tall, have dark hair, do have blue eyes, but am too lazy to shave more than once a week. In other words, my appearance, in the context of the match that evening, could be considered an ambiguous stimulus.

Near our hotel we passed a bar filled with locals who were getting ready for the match. They wore orange shirts and were clearly supporting our football team. We thought it safe to take a short look inside. Unlike the rest of our group, my ambiguous appearance did not fit the locals' notion of what typical Dutch people look like. As soon as they spotted me, several locals started to

loudly derogate me for being Italian. According to the locals I could put my pasta in a very special place.

How is it that I was so easily misperceived as an Italian person, within a split second? Up to now, I would love to believe that the only reason for being seen as Italian is that I dress a lot better than my colleagues. However, it probably had more to do with a fascinating process that kicked in once the locals perceived my face. It is called social categorization. It is such an efficient and tiny process, that we do not even notice it. It blends in with perception and helps us to understand the world around us. We now know that this process is the cognitive basis of all discrimination, prejudice, and stereotyping in the world. This dissertation explores that process in its earliest stages.

Contents

Chapter 1

Introduction

> "For the most part, we do not first see, and then define; we define first and then see." - Walter Lippmann (1922).

Perception involves an act of categorization (Bruner, 1957). We do not see a collection of geometrical figures. We see a chair. We understand what we see by relating it to what we already know, by applying knowledge about similar objects we encountered before to the object we perceive. "To categorize is to render discriminably different things equivalent, to group the objects and events and people around us into classes, and to respond to them in terms of their class membership rather than their uniqueness" (Bruner, Goodnow, & Austin, 1956). Categorization is an inevitable and necessary prerequisite for human cognitive functioning (Bruner, 1957; Rosch & Lloyd, 1978; E. E. Smith & Medin, 1981). We categorize other people just as we do other objects, in order to simplify our complex social environment (Allport, 1954; Lippmann, 1922; Tajfel, 1969). This is not a product of bad intentions, but of economy: our lim-

ited cognitive capacity does not enable us to treat every individual as unique (Fiske & Taylor, 1984). Instead, we automatically place people into gender, age, ethnic, or other categories (e.g., Brewer, 1988), in an immediate fashion without even noticing it (cf. Gibson, 1979; Neisser, 1987b). A perceived person just appears old or young, female or male to us without any effort on our part.

The field of social psychology has primarily focused its research effort on how categorization guides subsequent thought and behavior. However, for a long time it has neglected to investigate the pre-categorization phase in which appropriate categories are activated when a person is perceived. As a result, we know much less about how people categorize other people than we know about the consequences of categorizing other people. This dissertation is focused on the pre-categorization phase. In this introduction, we will first outline mainstream theories of social categorization. We will then discuss recent research on the pre-categorization phase, on the basis of which we present an extended model of social categorization. We will focus our discussion of the extended model on the categorization of faces, which are a major source of social information (Macrae & Quadflieg, 2010).

1.1 Social categorization

Categorical thinking is the foundation of prejudice and discrimination. As a result of categorization, people perceive members from the same category to be more similar and members from different categories to be more different than they really are (Tajfel, 1969). Information regarding the social category is generalized to a categorized person (Secord, Bevan, & Katz, 1956; Taylor, Fiske, Etcoff, & Ruderman, 1978). People experience category-based affect towards that person (Fiske, Neuberg, Beattie, & Milberg, 1987; Fiske & Pavelchak, 1986; Gre-

enwald, 1990). Furthermore, all subsequent information processing is guided by the activated social category (Brewer, 1988). For instance, people are more likely to remember category consistent information when they are under cognitive load, as is often the case in everyday life. On the other hand, when they are under low cognitive load, people seem to better remember category inconsistent information (Macrae, Hewstone, & Griffiths, 1993; Stangor & Duan, 1991), probably because they have enough resources to trigger attributional processes of inconsistency resolution (Sherman & Hamilton, 1994; Vonk & van Knippenberg, 1995; for an overview see Hilton & Von Hippel, 1996; Rojahn & Pettigrew, 1992; Stangor & McMillan, 1992). In both cases, memory for person information is influenced by categorization. Moreover, people primarily pay attention to category-relevant cues (Carver & Garza, 1984; Belmore & Hubbard, 1987), and seek out category-confirming information (Johnston & Macrae, 1994; Rothbart, 1981).

The cognitive structure that contains the information associated with a category is called the stereotype (Stangor & Schaller, 1996; van Knippenberg & Dijksterhuis, 2000). Stereotypes contain all knowledge, beliefs, and expectations about a category (Hamilton & Trolier, 1986; Hilton & Von Hippel, 1996). Upon categorization, stereotypes are almost always automatically activated (Devine, 1989, but see Gilbert & Hixon, 1991). Traditionally, stereotypes have been construed as cognitive schemas (Fiske & Linville, 1980; Taylor & Crocker, 1981), abstract knowledge structures consisting of propositions that specify the defining features and relevant attributes of a category.

Prejudice is most often construed as the negative attitude associated with a category (Dovidio & Gaertner, 2010; Fiske, 1998). Prejudice can be measured directly using explicit attitude self-report measures, or indirectly, using reaction time-based measurements such as the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998). The distinction between implicit (indirectly

measured) and explicit (directly measured) prejudice is important, because both types of prejudice predict different types of intergroup bias (Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Dovidio, Kawakami, Smoak, & Gaertner, 2008). Explicit prejudice generally correlates with deliberate forms of behavior, such as what a person says to an out-group member. Implicit prejudice, on the other hand, correlates with more subtle uncontrollable behavior, such as non-verbal behavior when speaking to out-group members (Dovidio, Kawakami, & Gaertner, 2002; McConnell & Leibold, 2001) or how much distance one maintains to out-group members (Dotsch & Wigboldus, 2008).

In sum, social categorization is the foundation of intergroup bias (Dijksterhuis, 2010; Dovidio & Gaertner, 2010). Figure 1.1 on page 5 depicts what might be called the mainstream model of categorical perception, compatible with most contemporary theories of person perception (Brewer, 1988; Fiske & Neuberg, 1990; also see Brewer & Feinstein, 1999). Upon perception, a person is immediately categorized, which automatically activates category related information (stereotype and prejudice). Subsequent thought and behavior are affected by the categorization and stereotype application.

1.2 Category allocation

Current models of social categorization, such as the one in Figure 1.1, describe primarily what Barsalou (1990) called the inference phase of categorization, the phase in which a perceiver goes beyond the information given based on category membership (Bruner, 1957). This reflects a strong research focus of social psychologists on the inference phase of categorization and contrasts with the focus of cognitive psychologists on what Barsalou called the access phase of categorization (e.g., Barsalou, 1983; Bruner, 1957; Medin & Smith, 1984; Mervis

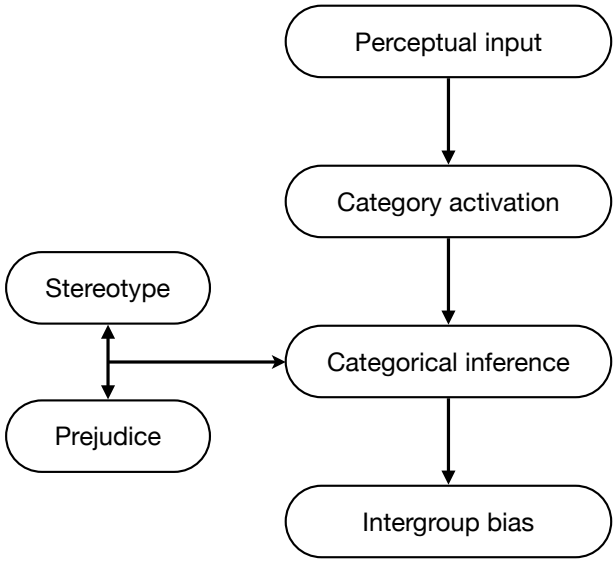

Figure 1.1: Mainstream model of categorical perception

& Rosch, 1981; Rosch & Lloyd, 1978; Rosch & Mervis, 1975). The access phase, or the category allocation phase as we will refer to it, is the stage of categorization in which the appropriate category is assigned to a perceived stimulus, and is represented in Figure 1.1 by the link between perception and categorization. This model, we will argue, is too simplistic. It is not that social psychologists have no theories about how categorization is involved in perception. Most implicitly assume that some kind of matching process goes on in which stimulus attributes are matched to category specifications, but this part of the process has never been explicated in mainstream theories of social categorization. However, the last two decades sparked a renewed interest in this specific part of the process, as will be discussed shortly.

The original lack of attention to the link between perception and categorization in social psychology is most obvious in Fiske and Neuberg's (1990) description of their influential continuum model of person perception. They simply stated that

> "the exact criteria for interpreting whether an instance fits a category are clearly important but cannot be our primary concern here" (p. 6).

A possible explanation of social psychology's disinterest in the category allocation phase up to the 90's is that in everyday life the social consequences of categorization have so much impact on society that originally all research effort had gone into investigating their pervasiveness. Another possible reason is of a more technological nature: presenting high quality graphical stimuli representing actual persons requires computer hardware which was not available until two decades ago. As a result, in most studies categories were activated by directly providing verbal category labels, thereby sidestepping completely the issue of how perception maps on categorization. Although categories in daily life are inferred from people's appearance, in the social psychological lab up to the 90's, people were generally replaced by category labels. Not surprisingly, Macrae and Quadflieg (2010) stated that

> "despite person perception constituting a central area of inquiry in social psychology for decades, two elements have surprisingly been absent in explorations of this topic - persons and perceptions" (p. 428).

As a result, an important stage of the categorization process, namely the transfer from perceptual representation to conceptual representation (e.g., Barsalou,

1999; Neisser, 1987a), is under-investigated in person perception literature. This dissertation aims to fill this gap by focusing specifically on the category allocation phase.

Because of the consistent use of verbal stimuli as opposed to perceptual stimuli, theories of categorical perception might be biased towards schematic representations. For some time, several researchers have tried to appeal to the field of person perception to stop relying solely on verbal stimuli to put the person and perception back in person perception. Zebrowitz and colleagues have repeatedly argued in favor of using more ecologically valid stimuli (Zebrowitz, 1996, 2006; McArthur & Baron, 1983; Zebrowitz & Montepare, 2008). Brewer (1988) has emphasized the importance of perceptual information in person perception, as have Carlston (1994), Feldman (1988), and McGarty (2002). In the last two decades the field has slowly shifted its research focus to more ecologically valid visual stimuli (e.g., Bijlstra, Holland, & Wigboldus, 2010; Blair, Judd, & Chapleau, 2004; Cloutier, Mason, & Macrae, 2005; Hugenberg & Bodenhausen, 2004; Hugenberg, 2005; Livingston & Brewer, 2002; Macrae, Bodenhausen, & Milne, 1995; Maddox & Gray, 2002; Quinn & Macrae, 2005). At the same time, research on the category allocation phase of person categorization has started to emerge (discussed below).

Capitalizing on these trends and building on research in the last two decades, in the current thesis we aim to present an extension to the mainstream model of social categorical perception that incorporates a more detailed description of the category allocation phase, during which perceptual input is interpreted conceptually. Moreover, we propose changes in the assumptions on the representational form and function of the stereotype construct, in order to deal with perceptual information. Many more extensions to the mainstream model are possible, but it is beyond the aim of the current thesis to provide a comprehensive model that encompasses all new developments in social categorization.

1.3 An extended model of social categorical perception

The purpose of the categorization process is to find the appropriate category for a perceived object. Categorization maps perceptual information onto concepts or categories (e.g., Neisser, 1987a). After categorization, visual input is no longer just a raw array of light intensity values, but consists of objects that have psychological meaning. How the categorization process accomplishes this for nonsocial, natural objects has been investigated for some time by cognitive psychologists (Barsalou, 1990). Whether knowledge about nonsocial object categorization can be generalized to person categorization has been a matter of considerable debate. Social psychologists have more than once argued that categorization of social stimuli might be fundamentally different from categorization of nonsocial objects. Persons, unlike objects, tend to be dynamic rather than static, active rather than passive, and to be perceived as causal agents (Lingle, Altom, & Medin, 1984; Ostrom, 1984). On the other hand, Brewer (1988) and Feldman (1988) have argued that such a distinction might be artificial and that there probably is no qualitative difference between social and nonsocial input in the early stages of the categorization process. Feldman proposed that any differences between social and nonsocial categorization might be explained by differences in expertise with the categories. He argued that the more experience someone has with a certain class of objects, the more elaborate the cognitive structure of the category. Because people interact with other people a lot, they build up great expertise with social categories. Similarly, bird spotters build up expertise with the category of birds. As a result the categorization of other people is just as 'special' as the categorization of birds for bird spotters. This is corroborated by recent evidence from neuroscience, showing that the fusiform face area (FFA), a brain area previously thought to be primarily active when perceiving social stimuli (faces), is just as active when perceiv-

ing objects of any nonsocial category (such as birds or cars), as long as people have built up visual expertise regarding that category (Kanwisher, 2000; Liu & Chaudhuri, 2003; Tarr & Gauthier, 2000; Xu, 2005). Even within the domain of faces, expertise matters. Experience with certain types of faces during the first year of development makes human infants more efficient in recognizing and discriminating those faces later in life (Scott & Monesson, 2009). Regardless of whether expertise solves the ostensible difference between social and nonsocial categorization, much can be learned from cognitive psychology, because "the disagreements and controversies [...] are not between cognitive and social psychologists [...] but between proponents of different approaches to categorization that span both social and cognitive psychology"(McGarty, 1999, p. 23).

We cannot claim full understanding of the categorization phenomenon "unless we also know who gets placed into which category and why" (Zebrowitz, 1996, p. 80). Specifically, the question of what social stimulus will end up in which category is central to this dissertation. Empirical work in the last two decades has significantly improved our understanding of the process of category selection. We now know that when multiple categories pertain to the same person, chronic and situational accessibility will be important determinants of which category is selected (Bruner, 1957; Oakes, 1987; van Knippenberg, van Twuyver, & Pepels, 1994). Accessibility refers to the ease with which a category comes to mind (Bruner, 1957; Higgins, 1989; van Knippenberg & Dijksterhuis, 2000). Some categories, such as gender or race categories, are chronically accessible (e.g., Stangor, Lynch, Duan, & Glas, 1992). Other categories, such as students from Nijmegen, can be made accessible by priming the respective categories (e.g., van Twuyver & van Knippenberg, 1995). Moreover, the activation of one category usually inhibits accessibility of other pertaining categories (Macrae et al., 1995; Bodenhausen & Macrae, 1998). Additionally, motivational processes may drive the selection of one category over another. For example, in an experiment by Sinclair and Kunda (1999), participants received feedback

from a Black doctor. When that feedback was positive, participants activated the doctor category. However, when the feedback from the same doctor was negative, they were more likely to activate the Black category. Furthermore, circumstances exist in which no social category will be activated at all, for instance when someone's processing goals do not require any social categorization (Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997).

Before any selection between accessible categories can be made, the cognitive system has to determine which categories pertain to a perceived person in the first place. It has been suggested that category pertinence depends on the normative fit between a perceived person and a category (Bruner, 1957; Oakes, 1987; van Knippenberg & Dijksterhuis, 2000). Normative fit refers to the match between stimulus characteristics and the characteristics a perceiver expects members of specific categories to have. In the cognitive literature, these expectations (also called category specifications) are thought to be part of the category structure itself. How fit with category specifications is established depends on the representational form of categories. Broadly, the cognitive literature can be divided in four conventional representational frameworks (E. E. Smith & Medin, 1981; also see McGarty, 1999). In the featural approach, in which categories are represented schematically, a stimulus fits a category when the stimulus has a critical number of category-relevant features. In the dimensional approach, category attributes vary on continuous dimensions. A stimulus can be represented as a point in multi-dimensional feature space and can be said to fit a category when the euclidian distance to a category center point is short enough, or falls within the multi-dimensional space spanned by the category. In the holistic approach, a stimulus fits a category when it matches a perceptual template for that category. In this sense, templates might be prototypical images of concrete objects. Finally, the exemplar approach holds that a category is represented by separate representations of its exemplars. There is fit between a category and a stimulus when that stimulus resembles stored category exem-

plars more than it resembles stored exemplars of other categories (e.g., Medin & Schaffer, 1978).

In social psychology, expected characteristics of group members are not part of a category representation itself, but part of the stereotype (Stangor & Schaller, 1996). Nonetheless, the described approaches to category structure can easily be mapped onto stereotype structure. For example, the featural approach is closest to the traditional view of stereotypes as schemata. The exemplar approach has been strongly advocated by Smith and colleagues (E. R. Smith & Zarate, 1990, 1992). However, the exact representational organization underlying categories and stereotypes is not the focus of this dissertation. Rather, we focus on the notion that whereas category specifications are part of the category in cognitive psychological theories, in social psychology category specifications are theoretically contained in stereotypes. As category specifications are critical for establishing fit, the implication is that stereotype content, which traditionally has been construed as affecting the inference phase after categorization takes place, might affect the category allocation phase too. This role has been hinted at, but has never been formally investigated. For instance, van Knippenberg and Dijksterhuis (2000) argued that "stereotypical expectancies are important determinants of spontaneous categorizations" (p. 113) and several of McGarty's (1999) constraint relations models acknowledged the influence of background knowledge on category membership decisions (p. 23). Finally, Blanz' (1999) model of category salience contains a clear link between stored stereotype content and perceived fit:

> "Stored stereotypes usually comprise rules to assign stimulus persons into aequivalent [sic] classes. That means stereotypes define the criteria for including or not including stimulus persons into a category" (pp. 45–46).

Unfortunately, Blanz' empirical work was aimed at keeping normative fit controlled and minimized, thereby leaving the relationship between stereotype content and fit largely unexplored (with Study 1 as notable exception, discussed shortly).

Importantly, there can be individual differences in stereotype content: While some might believe strongly that members of a social category are criminal, others might have weaker associations between the category and criminality. Because the stereotype contains the specifications against which category fit is established, people with slightly different stereotype content might put the same person in different categories under the same circumstances. This proposition lies at the heart of our extended model of social categorical perception and differs from mainstream views of categorization, as most researchers have come to think that "different people commonly categorize the same things in the same way" (McGarty, Yzerbyt, & Spears, 2002, p. 2). A study by Blanz (1999, Study 1) seems superficially related to our proposition. Blanz showed that participants who believed that a specific sports issue was more relevant for males than females, were more likely to use the gender dimension as a basis for categorization of people discussing this issue in a so-called 'Who Said What?' paradigm (Taylor et al., 1978). These results demonstrate that a certain categorization dimension can become more salient as a function of stereotype content, when multiple categorization dimensions are pertinent. However, our proposition is more subtle: given a certain salient categorization dimension, stereotype content might affect who gets placed in which category on that dimension.

Ultimately, this means the stereotype is an informational construct that supports two parts of categorical perception: pattern matching (by providing the information necessary to fit perceptual input to appropriate categories) and pattern completion (by providing the information to fill informational gaps

about the perceived person). Moreover, variation in stereotype content across individuals is predicted by their level of prejudice, that is, people who are prejudiced towards a particular group tend to associate the group stronger with negative stereotypical traits and weaker with positive stereotypical traits than less prejudiced individuals (Gordijn, Koomen, & Stapel, 2001; Wittenbrink, Judd, & Park, 1997). Because prejudice covaries with stereotype content and stereotypes affect fit, we propose that prejudice affects fit indirectly. Figure 1.2 depicts our extended model, including these predictions.

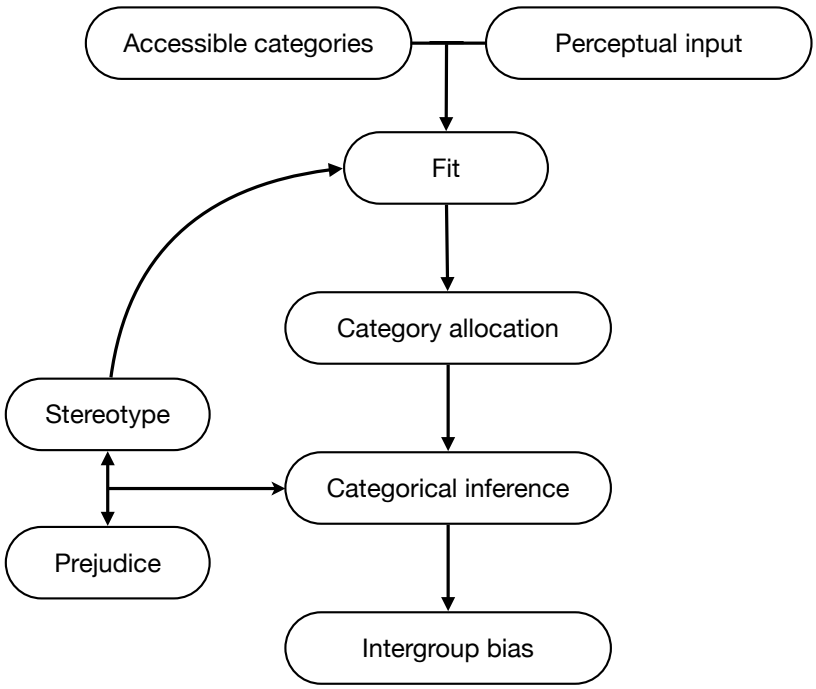

Figure 1.2: Extended model of categorical perception

In line with the appeal to the field to use more ecologically valid stimuli (e.g., Zebrowitz, 1996) and the fact that the effects of stereotypes and prejudice on the categorization process itself can only be tested when categorization of stimuli takes place (as opposed to using category labels as stimuli, thereby skipping the allocation phase of categorization), we chose to test this model in the domain of face perception.

1.4 Face perception

Faces are a major source of social information (Macrae & Quadflieg, 2010). They attract most of our attention when we perceive other people (Fletcher-Watson, Findlay, Leekam, & Benson, 2008). People infer person identity (e.g., Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000), emotional states (Ekman, 2003), personality traits (Hassin & Trope, 2000; Secord, Dukes, & Bevan, 1954; Todorov, Said, Engell, & Oosterhof, 2008; Zebrowitz & Montepare, 2008), physical attractiveness (Rhodes, 2006), and, importantly, category membership from faces (Brewer, 1988; Hugenberg & Sacco, 2008).

Similar to general person perception, face perception is affected by categorization and stereotyping. Categorizing a face as Black or White influences the perceived lightness of the face's skin (Levin & Banaji, 2006; see also Secord et al., 1956). Moreover, people are better at discriminating faces from two different categories than faces from the same category, even if the between-category physical differences are smaller than the within-category physical differences (e.g., Corneille, Hugenberg, & Potter, 2007; Levin & Beale, 2000). Categorization also biases face memory, such that faces are later remembered more prototypically (in the direction of the category prototypes) than they originally were (Corneille, Huart, Becquart, & Brédart, 2004; Huart, Corneille, &

Becquart, 2005). In line with automatic stereotype activation (Devine, 1989), categorized faces are attributed stereotypical traits (e.g., Ebner, 2008; Zebrowitz & Montepare, 1992) and emotions (e.g., Bijlstra et al., 2010; Hugenberg & Bodenhausen, 2003).

The present discussion of our extended model, however, focuses not on the consequences of face categorization, but on how face categorization is established: When does an accessible category pertain to a face? As argued above and in line with the extended model (and, e.g., Bruner, 1957; Oakes, 1987; van Knippenberg & Dijksterhuis, 2000) this is the case when there is normative fit. Because fit is established against category specifications contained in the stereotype, there should be better normative fit for faces that have a more stereotypical facial appearance. This raises the following important question: Do stereotypes contain representations of stereotypical faces?

We propose that stereotypes include a visual component which represents, among other things, typical faces. This is not a new proposition. In the beginning of the 20th century Lippmann (1922) originally defined stereotypes as "pictures in our heads", granting visual information precedence over semantic information. In line with the notion that not all information can easily be represented semantically (Klatzky, 1984), having a visual component makes it easier for stereotypes to represent visual information such as face configurations. In fact, the human brain has a complete neural network dedicated to processing and representing faces (Haxby et al., 2000; Ishai, 2008; but see Wiggett & Downing, 2008; Rossion, 2008). Therefore, the capacity for stereotypes to store what Klatzky called "pictoliteral" representations should not come as a surprise. Brewer (1988) made an extraordinarily strong case in favor of a visual component to stereotypes. As evidence, she cited research showing that category labels can (implicitly) elicit rich visual images of the kind of person represented by a category label (Ashmore, Del Boca, & Titus, 1984; Klatzky,

Martin, & Kane, 1982). Other research showed that physical appearances associated with gender stereotypes are highly linked to trait and role inferences, more so than any semantically represented component of stereotypes (Deaux & Lewis, 1984). Recent research on feature-based stereotyping also firmly establishes a direct link between stereotypical facial features and trait inferences or evaluative implicit judgments (e.g., Blair, Judd, Sadler, & Jenkins, 2002; Maddox & Gray, 2002; Maddox, 2004; Livingston & Brewer, 2002).

The hypothesis that stereotypes have visual representations is difficult to put to the test directly. Brewer (1988) deemed the question of whether stereotypes are in part represented visually empirically close to unresolvable (also see Anderson, 1978). The problem is that visual representations and semantic representations are difficult to disentangle in experiments, because often both can easily be re-described either verbally or visually. In other perceptual modalities such re-description might be more difficult. For instance, the auditive component of a stereotype might consist of what people expect typical group members' speech to sound like, in terms of e.g., typical pitch or accented speech (e.g., Gluszek & Dovidio, 2010). The olfactory component might store information about typical odors of group members, which could be difficult to label (for a discussion of labeling of odor and olfactory representation, see Wilson & Stevenson, 2003). Although the present dissertation is limited to the visual modality, in principle stereotypes might include representations of perceptual information in any modality. The advantage of a visual stereotype component (visual stereotype in short) is that in order to establish fit, the visual stereotype can be used as a template to match the incoming visual information to. This view is compatible with recent theorizing on general knowledge representation (e.g., Barsalou, 1999).

Typical faces represented in visual stereotypes can in our view contain information that is socially meaningful, beyond being just pure mathematical

averages of typical group members' faces. With socially meaningful information we mean facial features associated with personality traits. People efficiently infer traits from faces with great consensus (Oosterhof & Todorov, 2008; Hassin & Trope, 2000). The idea that facial features might represent personality traits dates from ancient Greece (McNeill, 1998; Zebrowitz, 1997) and was later formalized in the 18th century by Lavater (1772/1850) in "Essays on physiognomy". Because of his physiognomy, Charles Darwin was almost not allowed to make his now famous voyage on the Beagle to the Galapagos Islands; the ship's captain believed him to lack the necessary character as evidenced by the shape of his nose (Darwin, 1887/1950, p. 36). The 19th century Italian criminologist Cesare Lombroso claimed to have "scientifically" proven that criminal people could accurately be identified by certain facial characteristics (Lombroso, 1876/2006; also see Todorov, 2008; Zebrowitz, 1996). Even in the beginning of the 21th century, 80% of an Israeli student population believed that personality traits can accurately be inferred from faces (Hassin & Trope, 2000). Oosterhof and Todorov (2008) showed that people commonly infer personality from faces on the dimensions of trustworthiness and dominance. They can do so with high consensus even after just 100 ms presentation of the face (Willis & Todorov, 2006; also see Todorov, 2008). Sometimes, trait inferences from faces indeed correlate with traits or behaviors of the perceived persons. For instance, competence inferred from politicians faces is a good predictor of electoral outcome (Todorov, Mandisodza, Goren, & Hall, 2005). Likewise, inferences from faces of CEO's predict company profits (Rule & Ambady, 2008b). Importantly, Hassin and Trope showed that knowledge about a person's personality traits biases people's expectations about what this person's face will look like towards having facial features associated with those traits. From this finding, it is just a small step to postulating that knowledge about a group's personality traits biases people's expectations about what a typical group member's face will look like. Thus, we predict that the visual stereotype of typical group members' faces contains features that are associated with ste-

reotypical traits. Because people differ in terms of stereotype content, partly as a function of prejudice, we propose that people also differ in visual stereotype content as a function of prejudice. That is, facial visual stereotypes of more prejudiced individuals will contain more features associated with negative stereotypical traits (and less features associated with positive stereotypical traits) than those of less prejudiced individuals.

There is some evidence for our proposition in recent social psychological literature. Hugenberg and Bodenhausen (2004) found that the more implicitly prejudiced participants were, the more likely they were to categorize angry racially ambiguous faces as Black. Implicit prejudice was unrelated to the likelihood of categorizing happy faces as Black (also see Bijlstra et al., 2010; Bijlstra, Holland, Dotsch, & Wigboldus, in prep; Hutchings & Haddock, 2008). Assuming that anger is a stereotypical trait for the Black category, these data could be taken as evidence for an effect of stereotype content on categorization: more prejudiced people have a stronger belief that anger is an emotional expression that is stereotypically Black, and therefore are more likely to categorize angry faces as Black. Unfortunately, stereotype content was not measured in Hugenberg and Bodenhausen's work, rendering the evidence for our proposed influence of stereotype content on social categorization circumstantial at best. As we proposed that a visual mental representation mediates the effect of prejudice on categorization, in the present dissertation, we aimed to tap into this visual mental representation.

1.5 Measuring visual stereotype content

This dissertation will not provide a direct and definite test of visual stereotype representation, but tries to add to the accumulating body of evidence in favor

of such a view using novel data-driven methods to quantify visual stereotype content.[1] These particular methods are called psychophysical reverse correlation methods (Mangini & Biederman, 2004; Todorov et al., under review). These methods were originally developed in the domain of auditory cognition (Ahumada & Lovell, 1971), before they were used in research on vision (Ahumada, 1996, 2002; Beard & Ahumada, 1998; Solomon, 2002) and neurophysiology (Ringach & Shapley, 2004; Victor, 2005). The term *reverse* refers to a reversal of the statistical relationship between stimulus and response. In conventional paradigms, responses depend on meaningful manipulation of stimulus attributes. This relationship is quantified by correlating fixed stimulus attributes with responses. In reverse correlation paradigms, on the other hand, variations in stimulus attributes are random. The correlation between stimuli and responses can be used to model those variations in stimulus attributes that caused the acquired response pattern. In this type of analysis, the response variable is fixed whereas the stimulus attributes are random. This is the reverse of conventional analyses, hence the term: reverse correlation.

The particular class of reverse correlation methods used in this dissertation has been termed "superstitious perception" (Gosselin & Schyns, 2003; Mangini & Biederman, 2004), because they allow researchers to probe internal representations of categories without ever intentionally showing exemplars from those categories. We will refer to this kind of tasks as reverse correlation image classification tasks. Reverse correlation image classification methods are used when researchers want to examine participants' subjective internal representation of a category, without making any assumptions about what typical category members look like. In a typical reverse correlation task, participants classify variations of one single base face, unrelated to the categories of interest. Variations are created by distorting the base face with superimposed random

[1] This methodological part of the introduction is loosely based on a draft version of Todorov, Dotsch, Wigboldus, and Said (under review).

noise. Participants classify the noisy faces in whatever categories interest the researchers. Based on the classifications participants make, noise patterns can be generated which optimally visualize participants' subjective internal representation, called classification images.

Mangini and Biederman (2004, Study 3), in their validation of the reverse correlation image classification method, asked participants to make identity judgments. Specifically, the base image was a morph between the faces of John Travolta and Tom Cruise. Random sinusoid noise was superimposed on the morph to create variations (for examples of the resulting stimuli with different base images, see Figure 2.1 on page 24). Participants judged whether each variation was probably John Travolta, possibly John Travolta, possibly Tom Cruise, or probably Tom Cruise. Averaging all noise patterns classified as probably John Travolta resulted in a classification image showing what visual information yielded a John Travolta classification. Superimposing the classification image on the original base face resulted in an actual picture of John Travolta's face, or at the very least, an approximation of participants' subjective internal representation of his face. Likewise, averaging all noise patterns classified as probably Tom Cruise resulted in the Tom Cruise classification image. This is a good example of what can be achieved using a reverse correlation method. We adapted this method to tap into the visual component of group stereotypes.

1.6 The present dissertation

In the present dissertation we test our extended model of social categorical perception across three empirical chapters. Please note that there is not a one-on-one mapping of chapters to specific links in the model in Figure 1.2 on page 13. Chapter 2 focuses on the relation between prejudice and visual stereo-

types. We tested the prediction that real world visual stereotypes are affected by individual differences in prejudice in two studies. Dutch participants were asked to complete a reverse correlation task to assess what they expected typical Moroccan faces to look like. Moroccans are a highly stigmatized immigrant outgroup in the Netherlands (see Coenders, Lubbers, Scheepers, & Verkuyten, 2008) and are strongly associated with criminality (Gordijn et al., 2001). Afterwards, participants' implicit prejudice was measured. Independent participants rated the classification images resulting from the reverse correlation task on the traits *criminal* and *trustworthy*. More prejudiced participants generated more criminal and less trustworthy classification images than less prejudiced participants, indicating that indeed visual stereotypes are affected by prejudice, as predicted by our extended model.

The link between prejudice and visual stereotypes established in Chapter 2 is correlational. Therefore, in Chapter 3, we aimed to manipulate prejudice to examine the causal effect of prejudice on visual stereotypes using a stereotype formation paradigm (e.g., Crawford, Sherman, & Hamilton, 2002; Sherman, 1996; E. R. Smith & Zarate, 1990). Participants were asked to form an impression of a novel group, group X, based on exemplar faces and behavioral information. We manipulated behavioral information to be indicative of either the trait criminal or the trait trustworthy, while keeping exemplar faces constant. Afterwards, we assessed participants' visual stereotypes of group X using a reverse correlation task similar to the one used in Chapter 2. As in the studies in Chapter 2, the classification images were rated by independent participants on the traits criminal and trustworthy. Classification images of participants in the criminal group X condition were rated more criminal and less trustworthy than those of participants in the trustworthy group X condition. Because all participants saw the same exemplar faces during the stereotype formation task, these results identify group members' behavior as a cause of bias in visual stereotypes.

Whereas Chapters 2 and 3 focus on showing that mental representations of typical group members' faces can be biased, Chapter 4 investigates the consequences of biased visual stereotypes for social categorization. In this last empirical chapter we tested the predicted indirect influence of prejudice on category allocation: prejudice predicts variations in visual stereotype content (as shown in Chapters 2 and 3), which in turn should affect normative fit and subsequent category allocation. This was tested using categorization tasks.

Taken together, the three chapters provide first evidence for biases in visual stereotypes and subsequent categorization. In the final chapter, we discuss the extent to which our extended model of social categorical perception is supported by our empirical findings.

Chapter 2

Ethnic faces are biased in the prejudiced mind

Prejudice biases cognition, affect, and behavior toward outgroups (Fiske, 1998). We propose that prejudice also biases the way people conceptualize the facial appearance of outgroup members. Popular belief holds that people's personality traits are reflected in their facial features. Hence, people's beliefs about outgroup traits may also be reflected in what they think faces typical for particular outgroups look like. Because prejudiced people have more negatively stereotyped beliefs about outgroup characteristics (Fiske, 1998), we hypothesize that prejudiced people also have more negatively stereotyped mental representations of ethnic faces than less prejudiced people.

This chapter is based on Dotsch, Wigboldus, Langner, and van Knippenberg (2008).

To test this hypothesis we conducted two studies involving the category of Moroccans, a highly stigmatized immigrant group in the Netherlands. The first study consisted of two parts: image construction (Part 1) and image rating (Part 2). In Part 1, participants ($N = 28$) produced an image of a face based on their representation of Moroccan faces, using a forced-choice reverse-correlation image classification technique (Mangini & Biederman, 2004). In 390 trials participants repeatedly chose the most Moroccan-looking face from two stimulus faces presented side by side. All stimuli consisted of the same base face with random noise superimposed[1] (Figure 2.1). Within a single trial, one stimulus consisted of the base face with a random noise pattern added, and the other consisted of the base face with the same pattern subtracted. Averaging all stimulus faces a participant chose as most Moroccan results in a personal classification image (Figure 2.2). This classification image is a function of a participant's representation of Moroccan faces, the base face, and error.

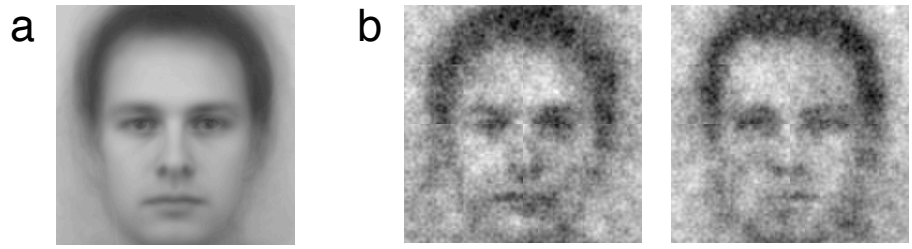

Figure 2.1: Base image used in all stimuli across both studies (a), examples of stimuli used in the studies (b)

[1]The base face was the neutral male mean of the Averaged Karolinska Directed Emotional Faces database (Lundqvist & Litton, 1998). The noise consisted of 60 superimposed sinusoid images: 6 orientations (0°, 30°, 60°, 90°, 120°, and 150°) x 5 spatial frequencies (1, 2, 4, 8, and 16 cycles per image) x 2 phases (0, $\pi/2$), with random contrasts.

The relative contribution of the base face to the resulting image is unknown, but the current technique seems to allow for a great variety of classification images given the same base face. To illustrate, Figure 2.2 shows the average Moroccan classification image (resulting from Part 1 of the Moroccan image study described above) and the average classification image of 30 independent participants who were instructed to select the most Chinese-looking stimulus using the same procedure.

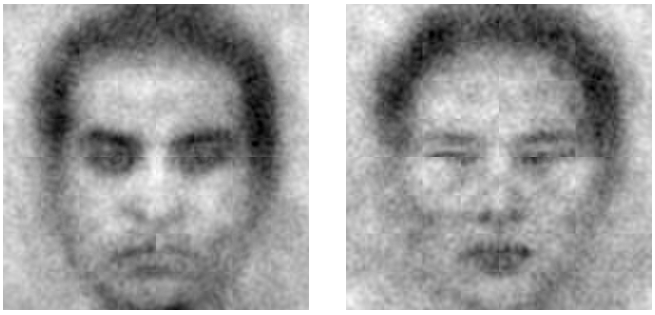

Figure 2.2: Averaged Moroccan classification image in Study 1 and averaged Chinese classification image of 30 independent participants

In Part 1 of the Moroccan image study, each participant's prejudice level was assessed using a Single Target Implicit Association Test (ST-IAT; see Greenwald et al., 1998; Bluemke & Friese, 2008). It measured the relative strength of negative associations compared to positive associations with Moroccan names as indicated by reaction times. Positive ST-IAT difference scores reflect stronger negative than positive associations with Moroccans ($M_{\text{difference}} = 34.77, SD = 74.92$). Based on this measure, participants were divided into low-, moderate-, and high-prejudice subgroups with cutoffs on the 33rd (3.97) and 66th percentile (66.66). For each subgroup, an average classification image was calculated representing this subgroup's average representation of a Moroccan face (Figure 2.3).

Figure 2.3: Classification images of the three subgroups (high, medium, and low prejudice) in Study 1. Scores below the images indicate trait ratings, on scales ranging from -3 (*not criminal* or *not trustworthy*) to 3 (*very criminal* or *very trustworthy*)

In Part 2, independent participants ($N = 70$) rated the three subgroup images on two traits related to the Moroccan stereotype. Participants were first shown the three subgroup images to familiarize them with the stimulus set. Subsequently, each of the three subgroup images was rated on two traits: *criminal* (stereotypical trait) and *trustworthy* (counter-stereotypical trait). Image order was counter-balanced and trait rating order was randomized. Analyses of variance on the criminality, $F(2, 68) = 39.00, p_{rep} > .99, \eta^2_{\text{partial}} = .53$, and trustworthiness ratings, $F(2, 68) = 43.95, p_{rep} > .99, \eta^2_{\text{partial}} = .56$, showed that the classification image of highly prejudiced participants was rated as more criminal ($p_{rep} > .99$) and less trustworthy ($p_{rep} > .99$) than the image of moderately prejudiced participants, which was rated as marginally more criminal ($p_{rep} = .85$) and less trustworthy ($p_{rep} = .95$) than the image of low prejudiced participants.

The results of the first study suggest that representations of Moroccan faces are biased by prejudice. We ran a second study using more trials in Part 1

(770 trials) but otherwise identical design to enhance the quality of individual participants' classification images ($N = 35$). This allowed us to replicate the findings of the first study on an individual rather than subgroup level. In Part 2, independent participants ($N = 55$) rated all individual classification images produced in Part 1 on criminality in one block and trustworthiness in another block. Block order was counterbalanced and image order within blocks was randomized. Within-participant standardized regression coefficients (βs), representing the relation between producers' prejudice level and the current participant's rating of the classification image, were calculated for each rater and each trait. Subsequent t-tests on the βs revealed that the more prejudiced the producers, the more criminal, $\bar{\beta} = .12, SD = .19, t(54) = 4.94, p_{rep} > .99$, and the less trustworthy, $\bar{\beta} = -.08, SD = .17, t(52) = 3.5, p_{rep} > .99$, their classification images were rated.

These results suggest that people's representations of ethnic faces are related to their level of prejudice. Future research should provide more insight into the nature of these representations. Furthermore, other factors that moderate mental representations of groups may be identified besides prejudice, such as context or specific derogatory group labels. The present results have important implications for whom people identify as members of stigmatized groups. They suggest that prejudiced people have a more criminal-looking prototype of Moroccan faces. Since more prototypical exemplars are processed more fluently (Winkielman, Halberstadt, Fazendeiro, & Catty, 2006), prejudiced individuals may find it easier to categorize criminal-looking Moroccan faces as Moroccan than innocent-looking Moroccan faces. This may function as a stereotype maintaining device.

Chapter 3

Visual stereotype formation: Behavioral information biases facial representations of new groups

Stereotypes are cognitive structures that contain the knowledge, beliefs, and expectations about groups (Hamilton & Trolier, 1986; Hilton & Von Hippel, 1996; Stangor & Schaller, 1996; van Knippenberg & Dijksterhuis, 2000). People form a great number of stereotypes in the course of their lives, through processes of (non-)conscious covariance detection (Ford & Stangor, 1992; Jussim, 1991; Lewicki, 1986), illusory correlation (Hamilton & Gifford, 1976; Hamilton

This chapter is based on Dotsch, Wigboldus, and van Knippenberg (under review).

& Sherman, 1989; Mullen & Johnson, 1990), or social learning and communication (e.g., Hirschfeld, 1995; Wigboldus, Semin, & Spears, 2000). These formation processes can be induced in lab experiments (e.g., Crawford et al., 2002; Sherman, 1996; E. R. Smith & Zarate, 1990). In these experiments, participants typically receive information about traits or behaviors of exemplars of a novel category. If those exemplars on average possess a certain trait and variability across exemplars is low, associations between groups and traits will be formed (Dijksterhuis & van Knippenberg, 1999). These stereotypical associations are formed more readily when group members are perceived as a unified entity (i.e., highly entitative; Campbell, 1958; Crawford et al., 2002; Hamilton & Sherman, 1996).

Most stereotype formation research assessed newly formed stereotype content at the level of verbal trait and behavior descriptions. On the basis of these measurements, one might assume stereotype content to be abstract and schematic (e.g., Higgins & Bargh, 1987). However, stereotypes were originally defined as "pictures in our heads" (Lippmann, 1922). McArthur and Baron (1983) emphasized the contribution of physical appearance to stereotyping, as did Brewer (1988), Carlston (1994), and McGarty (2002). Perceptual information may even be the primary basis of conceptual knowledge (Barsalou, 1999). In this view, the nature of stereotype content is for an important part perceptual (Zebrowitz, 1996). Perceptual stereotype content might contain visual information (e.g., group members' typical physical appearance) as well as information from any other modality. For instance, the auditive component of the stereotype might consist of what people expect typical group members' speech to sound like, in terms of e.g. typical pitch or accented speech (e.g., Gluszek & Dovidio, 2010).

Only recently have social psychologists begun to appreciate the visual component of stereotypes by using faces as stimuli in research (e.g., Bijlstra et al.,

2010; Blair et al., 2002; Hugenberg & Bodenhausen, 2004; Hugenberg, 2005; Livingston & Brewer, 2002; Macrae & Martin, 2007; Maddox & Gray, 2002; Mason, Cloutier, & Macrae, 2006). Indeed, in research on existing stereotypes "faces finally have found favor" (Zebrowitz, 2006). However, research on stereotype formation has, to our knowledge, as yet mostly ignored this visual component. With the current study, we aim to bridge that gap by assessing the visual component of newly formed stereotypes.

3.1 Faces as visual stereotype content

Because faces are an important source of information to establish category membership, group members' typical facial appearance is likely to be included in the visual stereotype. A newly formed visual stereotype might simply entail aggregating all previously encountered group members' facial configurations. On the other hand, faces are not just pure feature configurations, but take on social meaning as people infer traits from faces (Todorov et al., 2008; Zebrowitz & Montepare, 2008). Moreover, people use knowledge about someone's personality as a source of information about facial appearance (Hassin & Trope, 2000). Therefore, it is likely that after encountering several group members, the resulting visual stereotype is biased towards facial features associated with traits inferred from the behavior of those group members.

Visual content of stereotypes can be assessed with reverse correlation methods (Ahumada & Lovell, 1971; Dotsch et al., 2008; Mangini & Biederman, 2004). Specifically, these have been called superstitious perception methods (Gosselin & Schyns, 2003), because they allow researchers to probe internal representations of categories without ever intentionally showing exemplars from those categories. In a typical reverse correlation task, participants select from ran-

domly generated face variations those that best match their expectations of what typical category members look like. Based on their choices in a large number of trials, a classification image can be calculated, representing participants' visual stereotype content.

Using such a method, Dotsch et al. (2008) visualized what Dutch participants expected typical Moroccan faces to look like. Moroccans are a stigmatized immigrant minority group in the Netherlands (Coenders et al., 2008) and are strongly associated with the trait criminal (Gordijn et al., 2001). The visualized stereotype content in the Dotsch et al. study revealed that on average participants expected Moroccan faces to look criminal. Importantly, this effect was strongest for participants who had a more negative attitude towards Moroccans, i.e., who were more prejudiced. Thus, prejudiced people have visual stereotypes that are biased towards features associated with negative traits (also see Dotsch, Wigboldus, & van Knippenberg, under revision). However, the question remains whether these biased visual stereotypes can be caused by the inference of stereotypical traits from observed behaviors, as argued previously.

3.2 The current work

We propose the following causal relationship: information about behavior diagnostic of an out-group's traits will bias perceivers visual stereotypes of typical out-group faces towards having facial features corresponding with the inferred traits. We tested this hypothesis using a stereotype formation task in which we showed descriptions of exemplar behavior alongside exemplar faces. We manipulated exemplar behavioral information to reflect either trustworthy or criminal traits, while keeping exemplar faces constant. We then assessed participants' visual stereotype content using a reverse correlation task similar to

Dotsch et al. (2008). We predicted that the visualized stereotype would vary in line with the manipulation.

Moreover, in order to replicate the findings of Dotsch et al. (2008) that prejudice (i.e., a negative attitude) predicts bias in visual stereotype content, we measured participants' explicit and implicit attitude towards the novel group. Although implicit attitudes are often thought to stem from long-term socialization experiences (e.g., Rudman, 2004; Rudman, Phelan, & Heppen, 2007, but see Castelli, Carraro, Gawronski, & Gava, 2010), both explicit and implicit attitudes towards groups can be induced in the lab in just a single stereotype formation session (see Ratliff & Nosek, 2010). When no norm exists that prevents participants from explicitly stating a negative evaluation, there usually is a correlation between explicit and implicit attitudes (Greenwald, Poehlman, Uhlmann, & Banaji, 2009, also see Gawronski & Bodenhausen, 2006). We therefore expected both explicit and implicit attitudes to vary in line with the behavioral information manipulation. Moreover, we expected both implicit and explicit attitudes to mediate the effect of the manipulation on visual stereotype content.

3.3 Method

3.3.1 Participants

Seventy-seven students (13 males, $M_{\text{age}} = 22.36, SD = 4.91$) from Radboud University Nijmegen participated in the stereotype formation study. They received €5 or course credits for participating.

3.3.2 Overview and Design

The basic experiment consisted of two tasks, the stereotype formation and the reverse correlation task. In the stereotype formation task participants formed an impression of two novel groups, X and Y, based on exemplar faces and behavioral descriptions. For half of the participants the group X behaviors were predominantly indicative of the trait *trustworthy* and group Y behaviors predominantly indicative of the trait *criminal*. For the other half this was reversed. After a filler task, we visualized participants' representation of typical group X faces using a reverse correlation task (Dotsch et al., 2008; Mangini & Biederman, 2004), yielding so-called classification images. To assess the extent to which participants' representations of group X faces varied in line with the trustworthy or criminal behavior manipulation, independent participants rated the resulting classification images on both traits.

The experimental design consisted of one between-participants factor (group X behavioral information: trustworthy vs. criminal). Additionally, the design included several counter-balancing factors to control for, e.g., effects of order and exemplar faces. These will be explained in more detail later. After the reverse correlation task,we measured implicit and explicit attitude towards group X.

3.3.3 Materials

Stereotype formation task

In the stereotype formation task (adapted from Crawford et al., 2002) participants were instructed to form an impression of two novel groups, X and Y, based on exemplar faces and behavioral descriptions presented on a computer

screen. They received this information in two blocks, one block per group. Block order was counter-balanced. In total, participants viewed 20 exemplar faces and behaviors per group. In each trial a label of the current group (*Group X* or *Group Y*) was presented at the top of the screen, with an exemplar face in the center of the screen, and a behavioral description beneath it. The task was self-paced.

The exemplar faces consisted of one of two base faces (one for each group, counter-balanced across participants, see Figure 3.1 on page 36) with superimposed noise that distorted the faces to create subtle variations on the base face. The two base faces were selected from the Radboud Face Database (RaFD; Langner et al., 2010). The selected faces (models 7 and 23) were matched in valence and showed a neutral expression. The images were cropped to 512 x 512 pixels, converted to grayscale, and were blurred with a low-pass gaussian filter (with a kernel spanning 20 pixels in both image axes) to better match the spatial frequency band of the noise. The noise was constructed in the same manner as in Dotsch et al. (2008), i.e., by superimposing multiple layers of sinusoids with random amplitudes in six orientations, five spatial frequencies, and two phases[1]. We created two sets of 20 noise patterns, one set for each base face (counter-balanced across participants), resulting in 20 exemplar faces for one group and 20 exemplar faces for the other group (see Figure 3.1b on page 36 for examples).

Exemplar faces of one group were paired with 10 neutral behaviors (e.g., *This member of group X/Y crosses the street*) and 10 trustworthy behaviors (e.g., *This member of group X/Y returns the wallet he found*). Exemplar faces of the other group were paired with 10 different neutral behaviors and 10 criminal behavi-

[1]The random noise pattern consisted of superimposed truncated sinusoid images in 6 orientations (0°, 30°, 60°, 90°, 120°, and 150°) x 5 spatial frequencies (1, 2, 4, 8, and 16 cycles per image) x 2 phases (0, $\pi/2$), with random contrasts (amplitudes). In sum, the random noise was a function of 4092 parameters.

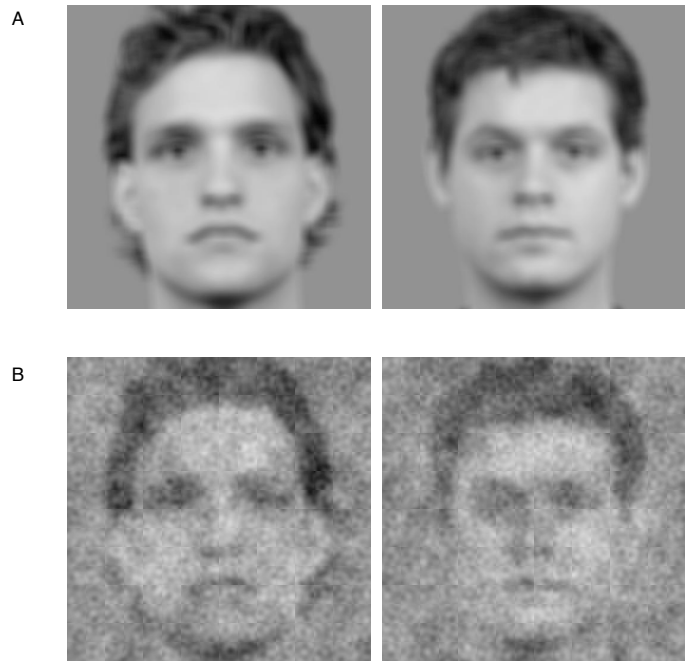

Figure 3.1: Base faces (a) and example stimuli in stereotype formation task for each base image (b)

ors (e.g., *This member of group X/Y robs another person in an alley*). The specific face-behavior pairings were randomized within groups. Which set of 10 neutral behaviors was associated with what group was counter-balanced across participants. Exemplar order was randomized.

Reverse correlation task

To visualize the face participants expected typical group X members to have, we used a two-alternative forced choice reverse correlation task adapted from Dotsch et al. (2008); also see Mangini and Biederman (2004). In 480 trials participants repeatedly chose from two faces presented side-by-side the face that most likely was a member of group X. All faces in this task were generated using as base face the average of the two blurred base faces from the stereotype formation task (Figure 3.2a). Averaging the two base faces ensured that stimuli in the reverse correlation task were equally likely to resemble group X and group Y faces. We generated random noise patterns in exactly the same manner as for the stereotype formation task (see Footnote 1). In a single trial, one stimulus consisted of the base face with a random noise pattern and the other with the inverse pattern superimposed (Figure 3.2b). Because the random noise patterns distorted the base face, the faces appeared to be different in every trial. All participants received identical sets of noisy stimuli, but in random order. Averaging all noise patterns chosen as group X member per participant resulted in individual classification images, representing what a participant expected typical group X members to look like. These classification images constitute an approximation of participants' visual group X stereotype, given the used base image, the set of random noise patterns, and error.

Because the reverse correlation task is a long and tedious task, it is not uncommon for participants to become demotivated after some time and start responding randomly. Trials on which participants responded faster than 300 ms were assumed to reflect no processing of the presented stimuli and therefore an indication of demotivation. To ensure high classification image quality, we decided a priori to exclude participants who responded faster than 300 ms on more than 25% of all trials.

Figure 3.2: Averaged base face (a), example stimuli with noise superimposed (b) and inverse noise superimposed (c)

Implicit Association Test

In the Implicit Association Test (IAT; Greenwald et al., 1998), participants classified positive and negative words (e.g., *love*, *death*) and the group labels *Group X* and *Group Y* into categories. Our IAT consisted of five blocks in the following fixed order:

1. A practice block (20 trials) in which participants classified positive words with the left key, and negative words with the right key;

2. A practice block (20 trials) in which participants classified positive words and the label *Group X* with the left key and negative words and the label *Group Y* with the right key;

3. A target block (40 trials), with the same key mapping as the previous block;

4. A practice block (20 trials), in which participants practiced the reversed mapping of the group labels: classifying positive words and the label

Group Y with the left key and negative words and the label *Group X* with the right key;

5. A target block (40 trials), with the same key mapping as the previous block.

In each block every type of stimulus was shown an equal number of times. Within blocks, the order of stimuli was randomized. Faster responses in Blocks 2 and 3 than in Blocks 4 and 5 were assumed to indicate stronger positive than negative associations with group X and stronger negative than positive associations with group Y.

Explicit measures

Using a 7-point scale ranging from 1 (*I strongly disagree*) to 7 (*I strongly agree*) we asked participants to indicate their agreement to two statements measuring how positive and how negative they evaluated group X (*I have a positive evaluation of group X*; *I have a negative evaluation of group X*).

Instructional Manipulation Check

To identify participants who did not take the experiment seriously we embedded an item in a bogus questionnaire with 7-point scales as instructional manipulation check (Oppenheimer, Meyvis, & Davidenko, 2009). The critical item read: "To confirm that you are reading the text on the screen, please respond by pressing 'p' ", and looked exactly like a normal item. Participants could respond by pressing 'p', indicating that they were paying attention, or with any

number as they would with a regular 7-point scale, indicating that they were not paying attention to the instructions on screen.

Rating task

To quantify the extent to which classification images reflected meaningful variation caused by the behavioral information manipulation 105 independent participants (25 males, $M_{\text{age}} = 22.48, SD = 4.29$) rated the individual classification images on trustworthiness and criminality. They received €2 or course credits in return. In a block per trait participants rated all images using a 9-point scale ranging from 1 (*not trustworthy/criminal*) to 9 (*very trustworthy/criminal*). Trials in which raters responded faster than 300 ms have been removed. For each classification image we calculated the averaged trustworthiness and criminality ratings. The average trustworthiness and criminality ratings correlated $r(58) = -.96$ with each other and are, hence, completely redundant. Therefore, only the average criminality ratings per image are reported as dependent variables in data analysis.

3.3.4 Procedure

Participants started with the stereotype formation task, in which they were asked to form an impression of group X and group Y based on group members' pictures and behavioral descriptions. It was explained to them that the behavioral descriptions represented typical behavior of the group member whose picture they were viewing and that random-noise was added to pictures to make it more difficult to recognize the faces underneath.

We aimed to create strong associations between groups and traits. Because transference of traits from individual group members to other members of the group is strongest for groups that are perceived to be highly entitative (e.g., Crawford et al., 2002), all participants received the same high entitativity instruction before each block of the stereotype formation task:

> "The members of group X [Y] are very similar to each other and do not differ in many ways from each other. The members come from similar backgrounds and have the same opinions, similar important beliefs, and similar personalities. Across a variety of situations, members of group X [Y] will act in a similar manner." (Crawford et al., 2002, p. 1080).

After the stereotype formation task, participants completed an unrelated filler task, which lasted on average about 18 minutes. They then proceeded with the reverse correlation task, the IAT, explicit measures, and instructional manipulation check. Afterwards, participants were debriefed and thanked.

3.4 Results

3.4.1 Data preprocessing

Implicit Association Test

We processed the IAT data using the classic algorithm (Greenwald et al., 1998). The first two trials of each block and incorrect trials (6.23%) were removed from

analysis. Latencies below 300 ms were recoded to 300 ms ($< 1\%$). Latencies above 3000 ms were recoded to 3000 ms ($< 1\%$). All latencies were then log-transformed (untransformed latencies are reported). We calculated IAT scores for each participant by subtracting the average response latency in Blocks 2 and 3 from the average response latency in Blocks 4 and 5. A positive IAT score therefore indicated stronger positive than negative associations with group X (and stronger negative than positive associations with group Y).

Reverse correlation task

We generated classification images representing what participants expected typical group X members to look like (i.e., their visual group X stereotype) based on the reverse correlation data. Trials with response latencies below 300 ms were excluded from analysis (2.20%). The classification images were calculated by averaging the sinusoid amplitude parameters of the selected noise patterns. The resulting average noise patterns were then superimposed on the base image to create the final classification image for each participant. In order to visually illustrate the effect of the behavioral information manipulation we calculated averaged classification images for the participants in the criminal and trustworthy conditions separately (see Figure 3.3 on page 43). The individual classification images were rated on criminality by independent participants as described in the method section. For each image we calculated the averaged criminality rating, which was used as dependent variable in subsequent analyses.

Figure 3.3: Averaged classification images for trustworthy behavioral information condition (a) and for criminal behavioral information condition (b)

Exclusion criteria

To ensure good classification image quality, 15 participants who in the reverse correlation task responded faster than 300 ms on more than 25% of the trials were assumed to be demotivated by the large number of trials. They were therefore removed from the data on a priori grounds, as was described in the method section and their classification images were not included in the rating task. One participant did not give the correct answer on the instructional manipulation check and was removed from further analysis. Three participants who scored 2.5 *SD* below or above the average IAT score within their respective behavioral information conditions were removed from analysis. The final data set contained 58 participants.

3.4.2 Analyses

Preliminary analyses indicated that none of the counterbalancing factors from the stereotype formation task interacted with the effects of interest on crimin-

ality ratings (all *F*s < 1). We therefore dropped the counterbalancing factors from subsequent analyses, leaving only one between-subjects factor: behavioral information condition (criminal vs. trustworthy). As predicted, classification images of participants in the criminal behavioral information condition were rated as more criminal ($M = 5.26, SD = 0.70, n = 31$) than classification images of participants in the trustworthy behavioral information condition ($M = 4.86, SD = 0.77, n = 27$), $t(56) = 2.07, p = .04, d = 0.55$.

Participants' IAT scores also varied as a function of behavioral information condition, $t(56) = 2.16, p = .04, d = 0.58$. Participants in the trustworthy behavioral information condition had stronger positive than negative associations with group X ($M = 65.02, SD = 75.07$) compared to participants in the criminal behavioral information condition ($M = 23.59, SD = 67.34$). Moreover, participants in the trustworthy behavioral information condition had more positive ($M = 6.26, SD = 0.94$ vs. $M = 2.77, SD = 1.89$) and less negative ($M = 2.56, SD = 1.65$ vs. $M = 5.32, SD = 1.90$) explicit evaluations of Group X than those in the criminal behavioral information condition, respectively $t(56) = 8.67, p < .01, d = 2.32$, and $t(56) = 5.52, p < .01, d = 1.48$. Positive and negative explicit evaluations were highly correlated with each other ($r = -.79$). Therefore, in subsequent analyses the difference score (positive $-$ negative evaluation) was used as a general measure of explicit evaluation. A higher difference score reflects a more positive explicit evaluation.

We predicted that both implicit and explicit attitudes would mediate the behavioral information effect on classification image criminality. These hypotheses can be tested using within rater regressions to capture the mediation for each rater (i.e., policy capturing; Aarts, Verplanken, & van Knippenberg, 1997; Brehmer & Joyce, 1988; Dotsch et al., 2008). In this analytical strategy, the hypotheses are confirmed when we demonstrate on average a relation between behavioral information manipulation and classification image criminality per

rater (i.e., $\bar{\beta}_1$), a relation between the potential mediator (IAT scores or explicit evaluation) and classification image criminality per rater (i.e., $\bar{\beta}_2$), and a relation between behavioral information manipulation and classification image criminality per rater when controlling for the mediator (i.e., $\bar{\beta}_3$) which should be weakened (i.e., $\bar{\beta}_3 < \bar{\beta}_1$). Note that a relation between behavioral information manipulation and the potential mediators (IAT scores and explicit evaluation) has already been established above. As can be seen in Table 3.1 on page 46, which summarizes the within rater regressions, the effect of behavioral information manipulation on classification image criminality ratings was partially mediated by IAT scores (as $\bar{\beta}_3$ was different from 1, $p < .01$, but also different from zero, $p < .01$) and fully mediated by explicit evaluation (as $\bar{\beta}_3$ was different from 1, $p < .01$, and not different from zero, $n.s.$). The reverse model, that the effect of behavioral information manipulation on IAT scores is partially mediated by classification image criminality, was also significant.

Table 3.2 on page 47 reports the entire correlation matrix on the level of participants (instead of on the level of raters as was the case in Table 1). The first column of Table 3.2 shows that the manipulation correlates stronger with explicit evaluation than with IAT scores ($z = 3.14, p < .01$). It is possible that the IAT scores contain variance that uniquely predicts classification image criminality variance above and beyond the explicit evaluation. However, when controlling for explicit evaluation while regressing IAT scores on classification image criminality ratings using the policy capturing method described above, the IAT $\bar{\beta}$ was not significantly different from zero. Moreover, explicit attitudes predicted classification image criminality when controlling for IAT scores, $\bar{\beta} = -.20, t(104) = 12.82, p < .01$.

Table 3.1: *Policy capturing analyses of mediation of the effect of behavioral information manipulation on classification image criminality ratings*

		Mediator					
		Implicit			Explicit		
	Predictor	*M*		*SD*	*M*		*SD*
β_1	Effect of behavioral information on criminality ratings	.13	**	.13	.13	**	.13
β_2	Effect of mediator on criminality ratings	.06	**	.10	-.19	**	.15
β_3	Effect of behavioral information on criminality ratings when controlling for mediator	.12	**	.13	.00		.16
	Test of mediation	*t*		*p*	*t*		*p*
	$\bar{\beta}_3 < \bar{\beta}_1$	3.34		.01	10.43		.01

* significantly different from 0, $p < .05$; ** significantly different from 0, $p < .01$; all *df*s = 104.

3.5 Discussion

Using a stereotype formation task we tested whether behavioral information diagnostic of an out-group's traits biases perceivers' visual stereotypes of typical out-group faces towards having facial features corresponding with the inferred traits. Our hypothesis was confirmed: visual stereotypes of participants in the criminal behavioral information condition were judged to be more criminal than those of participants in the trustworthy behavioral information condition. Because the presented exemplar faces were kept constant, the reported bias in visual stereotypes could only be caused by exemplars' behavioral in-

Table 3.2: *Correlation matrix based on between-participants covariance*

	1.	2.	3.
1. Manipulation (1 = criminal, -1 = trustworthy)			
2. Classification Image Criminality	.27		
3. IAT Score (Higher is more positive)	-.28	-.27	
4. Explicit Evaluation (Higher is more positive)	-.71	-.49	.35

All correlations $p < .05$.

formation. To our knowledge, the current work is the first in demonstrating experimental effects of behavioral information on visual stereotypes. These findings suggest that in every day life people indeed have biased visual representations of social groups, as Lippmann (1922) suspected a long time ago.

Participants' explicit and implicit attitudes were in line with the behavioral information manipulation. This replicates work by Ratliff and Nosek (2010), who induced both implicit and explicit attitudes in a similar stereotype formation paradigm, using personality traits instead of behavioral information, and without any visual information. The behavioral information manipulation had a stronger effect on explicit attitudes than implicit attitudes. This fits theories stating that the implicit system learns more slowly than the explicit system (Rydell & McConnell, 2006; Sloman, 1996; E. R. Smith & DeCoster, 2000). Perhaps leaving more time between stereotype formation and attitude assessment or presenting more group exemplars would increase the effect of behavioral information on implicit attitudes. Additionally, the IAT might be a noisier measure than explicit ratings. For instance, the IAT score not only reflects implicit attitudes, but also individual differences in task switching capacity (Mierke & Klauer, 2003; Klauer & Mierke, 2005), generating spread across participants unrelated to the behavioral information manipulation.

Moreover, the IAT score showed a positivity bias (i.e., the average IAT score was not negative in the criminal behavioral information condition). One possible source of bias is the fixed IAT block order. Group X was always first paired with positive stimuli (and Y with negative). This may have slightly strengthened the association of group X with positive valence (and Y with negative) as a result of which the later reversal (i.e., paring group X with negative and Y with positive) may have yielded slower responses. Thus, this order of IAT blocks may have artifactually produced slightly more positive evaluations of group X and slightly more negative evaluations of group Y. Another source of bias might be the fact that the IAT, pitting group X against group Y, was administered after participants performed the extensive reverse correlation task which was focused only on group X. This greater exposure to group X than to group Y might have induced a positivity bias towards group X in a similar way as the implicit partisanship effect (Greenwald, Pickrell, & Farnham, 2002; Pinter & Greenwald, 2005). Nonetheless, people had relatively more negative implicit attitudes in the criminal behavioral information condition than in the trustworthy behavioral information condition.

Participants' implicit attitudes mediated the effect of the behavioral information manipulation on visual stereotypes, such that participants in the criminal behavioral information manipulation had more negative implicit attitudes, which in turn was related to a more criminal-looking visual stereotype. This essentially is a replication of Dotsch et al. (2008), in which more implicitly prejudiced participants (i.e., participants with stronger negative implicit attitudes) had more criminal-looking mental representations of typical out-group faces. However, behavioral information might not be the only source of the biased implicit attitudes and face representations. Because people read trait information from faces (e.g., Todorov et al., 2008), exemplar facial appearance might further bias attitudes and face representations, perhaps in interaction with behavioral information. This might be addressed in future research. Note that in

the present study exemplar facial appearance was kept constant across behavioral information conditions.

Explicit attitudes likewise mediated the effect of the behavioral information manipulation on visual stereotypes. Here the alternative mediational model of behavioral information affecting explicit attitudes through the visual stereotype was rejected on statistical grounds. Dotsch et al. (2008) did not report any correlations between explicit attitudes and visual stereotypes of Moroccan faces. Re-analysis of their data (Dotsch et al., Study 2) for the purpose of the current work showed that neither positive nor negative explicit attitudes correlated with the criminality or trustworthiness judgments of participants' visual stereotypes of Moroccan faces. This might be explained by a social norm not to appear explicitly prejudiced towards real-world groups (Crandall, Eshleman, & O'Brien, 2002; Dovidio & Gaertner, 2004). This norm might not be in place for the novel group in the current study.

It should be noted that one alternative mediational model, entailing that behavioral information condition affected implicit attitudes through the visual stereotype, cannot be excluded from a statistical point of view. This would mean that behavioral information affects explicit attitudes, which affect the visual stereotype, which in turn affects implicit attitudes. This model is theoretically less parsimonious than the simple model we proposed, i.e., that behavioral information affects the visual stereotype through both implicit and explicit attitudes. Importantly, evidence for the direct link between information manipulation and implicit attitude is provided by the previously mentioned work by Ratliff and Nosek (2010). The critical test in this respect would be to manipulate the mediators independently with a stronger induction of implicit attitudes (see e.g., McConnell, Rydell, Strain, & Mackie, 2008).

In the stereotype formation task, each group was assigned only one trait, opposite in valence. It is therefore unclear whether the classification images reflect valence or actually the traits criminal or trustworthy. Future research could tease apart the influence of valence and specific traits (similar to Amodio & Devine, 2006) by inducing multidimensional stereotypes.

The current work demonstrates the usefulness of reverse correlation tasks to lay bare the visual content of stereotypes. However, the outcome of reverse correlation tasks, a classification image, should not be equated with visual stereotype content, however tempting. At most, a classification image can be argued to be an approximation of visual stereotype content, affected and constrained by contextual and task-specific factors, such as the underlying base face image, the specific set of noise patterns, and motivation. This is not unlike other measurements in psychological research, which are approximations of a not directly observable latent construct (Lord & Novick, 1968).

The major contribution of the present study is that it establishes a causal link between behavioral information about exemplars and the expected facial appearance of group members. With this study, we hope to entice the field of stereotype formation to shift its focus on abstract and schematic stereotype content to more perceptual stereotype content.

Chapter 4

Biased allocation of faces to social categories

As much as we would like to individuate everyone we meet, we effortlessly and automatically categorize persons into groups to simplify and make sense of the enormous amount of social information in the world (Allport, 1954). Although much research exists on the consequences of categorization (e.g., Fiske, 1998; Tajfel, 1969; Turner, 1987), our knowledge about the category selection process itself is rather limited. We cannot claim full understanding of the categorization phenomenon "unless we also know who gets placed into what categories and why" (Zebrowitz, 1996, p. 80). Processing goals, motivational states, and (chronic) accessibility have been identified as some of the determinants of category selection and paint a clear picture of why specific categories have a higher probability to be selected when a target person belongs to mul-

This chapter is based on Dotsch, Wigboldus, and van Knippenberg (under revision).

tiple categories (Hugenberg & Sacco, 2008; Macrae, Bodenhausen, Milne, & Calvini, 1999; Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997; van Knippenberg, van Twuyver, & Pepels, 1994). However, before this selection can be made, the cognitive system has to determine which categories pertain to a perceived person in the first place. In the current paper we test the prediction that at this early stage of social categorization, the very process of category allocation is biased.

Category allocation has been investigated as early as the late 1950's. Pettigrew, Allport, and Barnett (1958) conducted a binocular rivalry experiment in South Africa in which they briefly presented, among other combinations, a White face to one eye and a face with darker skin tone (colored, African, or Indian) to the other eye. Despite the dual stimulus, participants perceive this as one single face. Afrikaners, i.e., white people who in the past were generally in favor of racial separation (Dubow, 1992), were more prone to allocate any combination of White and non-white stimulus faces to a non-White category than English White, African, or Indian participants. Pettigrew et al. explained these findings as an effect of prejudice on the allocation of ambiguous exemplars to racial categories. These data constitute preliminary evidence for biases in category allocation.

When does a category pertain to a perceived person? It has been suggested that there has to be normative fit between that person and the category (Bruner, 1957; Oakes, 1987; van Knippenberg & Dijksterhuis, 2000). Normative fit refers to the match between stimulus characteristics and the characteristics a perceiver expects members of specific categories to have. For instance, when the features or configuration of a perceived face matches someone's expectations of what typical Chinese faces look like, there is normative fit. The stereotype is the cognitive structure containing these expectations (Stangor & Schaller, 1996; van Knippenberg & Dijksterhuis, 2000). Importantly, this means

that normative fit is affected by the idiosyncratic content of the stereotype. If a culturally shared stereotype holds that members of a specific category are criminal, a person endorsing this stereotype will expect category members to look criminal, and therefore, normative fit will be enhanced when this person perceives a criminal-looking person.

There can be individual differences in the extent to which people endorse culturally shared stereotypes: While some might believe strongly that members of a social category are criminal, others might have weaker associations between the category and criminality. Importantly, this variation across individuals is predicted by their level of prejudice, i.e., the extent to which they have a negative evaluation of the category as a whole (Gordijn et al., 2001; Wittenbrink et al., 1997). To the extent that highly prejudiced people believe category members to be criminal, they will expect people of that group to look more criminal. This relationship recently has been established within the domain of faces. Dotsch, Wigboldus, Langner, and van Knippenberg (2008), used a forced-choice paradigm to reconstruct what people believed typical Moroccan faces looked like (Moroccans are a highly stigmatized immigrant group in the Netherlands, see Coenders et al., 2008; Verkuyten & Zaremba, 2005, and strongly associated with the trait criminal; Gordijn et al.). Facial images produced by highly prejudiced people depicted faces that were more criminal-looking than those produced by people low in prejudice.

Because highly prejudiced individuals expect Moroccan people to be more criminal, criminal-looking faces have better normative fit for the Moroccan category and, therefore, highly prejudiced individuals should be more likely to categorize criminal-looking stimulus faces as Moroccan.

Alternatively, Ruys, Dijksterhuis, and Corneille (2008) proposed an evaluative fit hypothesis such that social categorization is facilitated for exemplars that are

evaluatively congruent with the target category. With regard to the Moroccan category this means that because highly prejudiced individuals evaluate the Moroccan category more negatively, negative-looking faces have better evaluative fit for the Moroccan category and, therefore, highly prejudiced individuals should be more likely to categorize criminal-looking stimulus faces as Moroccan. The difference between evaluative fit and normative fit is that the first only concerns valence congruity due to prejudice, while the latter involves stereotypicality. It is often difficult to disentangle prejudice-related valence from stereotyping (e.g., Amodio & Devine, 2006; Wittenbrink et al., 1997), because for the majority of social categories the two concepts seem to be confounded: e.g., the trait criminal – which is stereotypical for Moroccans – is negative, and the Moroccan category itself is negatively evaluated. However, in our view prejudice and stereotyping have clearly identifiable – but interacting – contributions to fit.

Evaluative fit cannot possibly account for the vast number of potential social categories that can be allocated given a stimulus of some valence. The unidimensional concept of valence in the majority of cases is not predictive for category membership: not all negatively evaluated people are members of one specific negative category. Fit by necessity depends additionally on a concept of higher dimensionality: the stereotype. The stereotype contains those traits that are predictive of group membership (see Le Pelley et al., 2010). Some of those traits might have the same valence as the category, others might not. Therefore, in our view, faces with features associated with traits that are predictive of group membership (i.e., stereotypical traits) will have enhanced fit, regardless of valence. If a perceiver has a negative evaluation of that category (i.e., is prejudiced), faces with features that are both stereotypical and negative will have enhanced fit.

In sum, we propose that prejudice enhances fit for faces with negative features, only to the extent that these features are associated with stereotypical traits. This hypothesis differs from the evaluative fit hypothesis in that prejudice does not enhance fit for faces with features associated with any negative trait, but does so on the condition that this negatively valenced trait is stereotypical. In three studies, the present paper aims to test this normative fit hypothesis. Study 1 aims to show that people indeed over-allocate faces with features associated with negative stereotype-relevant traits to a negatively evaluated category[1] and that this happens to a greater extent for more prejudiced people. In Study 2 we aim to show that only stereotype-relevant negative traits, but not stereotype-irrelevant negative traits elicit over-allocation. Finally, in Study 3 we aim to show that even valence-incongruent traits may elicit over-allocation under the condition that these traits are stereotype-relevant, but then prejudiced participants will not display enhanced over-allocation.

4.1 Study 1

Study 1 was designed to show that more prejudiced people allocate more faces with features associated with a negative stereotype-relevant trait (criminal) to a stigmatized category (Moroccans). Dutch participants were instructed to categorize faces as Moroccan or non-Moroccan. Half of the faces were manipulated to have more criminal-looking features. After the categorization task implicit prejudice was measured. Dutch people associate Moroccans strongly with the trait criminal (Gordijn et al., 2001), thus, the trait is highly stereotype-relevant. Because normative fit is enhanced for faces with features associated

[1] By the over-allocation of faces with features associated with stereotype-relevant traits we mean that participants allocate more faces with features associated with stereotype-relevant traits to the target category than faces without those features.

with stereotype-relevant traits, we expected participants to allocate a larger percentage of the criminal-looking faces to the Moroccan category than the non-manipulated faces. Moreover, because the trait criminal is both stereotype-relevant and negatively valenced, we expected more prejudiced participants to show this effect to a greater extent.

4.1.1 Method

Participants

Nineteen male and 73 female Dutch-speaking students of the Radboud University Nijmegen participated in this study ($M_{\text{age}} = 21.59, SD = 2.78$). In return, participants received course credit or €4.

Design

The experiment employed a mixed model design with face set (Moroccan-looking vs. criminal Moroccan-looking) as within-subjects variable and implicit prejudice as continuous between-subjects variable. The dependent variable was the percentage of faces categorized as Moroccan.

Materials

Criminal features In order to manipulate criminal-looking facial features, a pilot study was run to construct noisy images of what people thought a typical criminal face looks like. In the pilot, 13 male and 22 female students of the

Radboud University Nijmegen ($M_{\text{age}} = 21.17, SD = 2.74$) completed a forced-choice version of a reverse-correlation image classification task (Dotsch et al., 2008; Mangini & Biederman, 2004). In this task participants were repeatedly presented with two stimuli side by side (see Figure 4.1b on page 58). Each stimulus consisted of two random noise patterns superimposed over a base image (see Figure 4.1a). The base image in all reverse correlation tasks throughout this paper was the neutral male face of the Averaged Karolinska Directed Emotional Faces Database (Lundqvist & Litton, 1998). The noise patterns were randomly generated at every trial[2]. Within a single trial, one stimulus consisted of the base face with a random-noise pattern and the other with the inverse pattern superimposed. Because the random-noise patterns distorted the base face, the faces appeared to be different in every trial. Participants were instructed to select the most criminal-looking face over 770 trials. Averaging all noise patterns selected as most criminal-looking across all participants resulted in a classification image representing what our pilot participants thought a typical criminal face looked like (see Figure 4.2b on page 59).

Stimuli Two sets of stimuli were used in Study 1: 35 Moroccan-looking and 35 criminal Moroccan-looking faces. The Moroccan-looking faces were the classification images of the individual participants in the Dotsch et al. (2008, Study 2) experiment (see Figure 4.2a for the average Moroccan classification image and Figure 4.2c for an example stimulus). These classification images were the result of the same procedure as the pilot study described above, but with the instruction to select the most Moroccan-looking face. The criminal Moroccan-looking faces used in the current study were generated by superimposing the average criminal classification image resulting from the pilot

[2]The random noise pattern consisted of superimposed truncated sinusoid images in 6 orientations (0°, 30°, 60°, 90°, 120°, and 150°) x 5 spatial frequencies (1, 2, 4, 8, and 16 cycles per image) x 2 phases (0, $\pi/2$), with random contrasts (amplitudes). In sum, the random noise was a function of 4092 parameters.

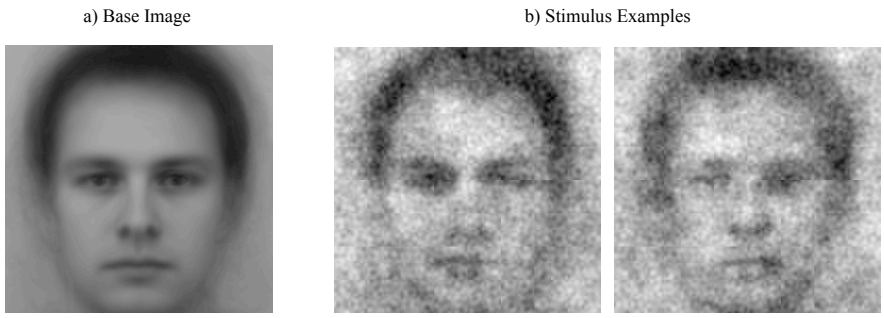

Figure 4.1: Base image (a) and example of stimuli (b) used in a trial of pilot study

study on each individual Moroccan-looking face from the Dotsch et al. study (see Figure 4.2d). As can be seen in Figure 4.2, the differences between the Moroccan-looking faces and criminal Moroccan-looking faces are subtle and difficult to perceive when presented side by side. A video clip demonstrating the differences more clearly with faces presented in succession is available at http://web.me.com/rdotsch/onlinematerials/stimuli.mov.

Procedure

Categorization task Participants first completed a categorization task in which the 35 Moroccan-looking and 35 criminal Moroccan-looking faces were presented on a computer screen one by one in random order. Each face was presented twice, resulting in a total of 140 trials. Participants were asked to categorize each face as either Moroccan or non-Moroccan as quickly as possible by pressing one of two category keys. They were told that noise was added to the faces to increase task difficulty. Participants were asked to respond as soon as a

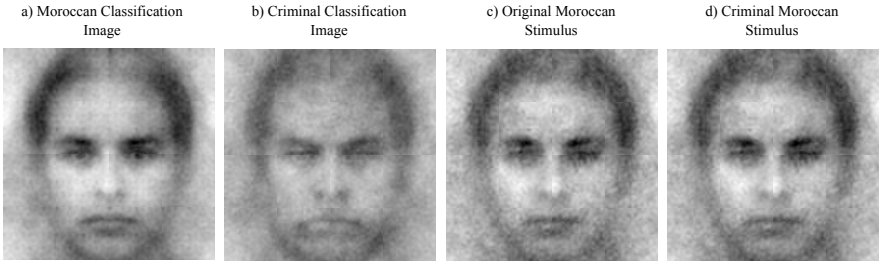

Figure 4.2: Moroccan classification image (a), criminal classification image (b), and example stimuli used in study 1 (c, d)

face appeared on the screen. Faces were presented until a response was given. There was no response-window. The inter-trial interval was 500 ms.

Implicit prejudice Participants then moved on to a Single Target Implicit Association Test (ST-IAT; Bluemke & Friese, 2008; Dotsch & Wigboldus, 2008). This task measured indirectly how strongly participants associated Moroccan names (e.g., Ibrahim, Rachid) with positive and negative words (e.g., love, peace, war, pain). The ST-IAT consisted of a practice, compatible, and incompatible block. In the practice block participants classified 10 positive words with one key and 10 negative words with another. In the compatible block participants classified 20 positive words with one key and 10 negative words and 10 Moroccan names with another. In the incompatible block, participants classified 10 positive words and 10 Moroccan names with one key and 20 negative words with another. Block order was counter-balanced. Within blocks, stimuli were presented in random order. When participants made an incorrect classification, error feedback was presented for 1000 ms. Shorter response latencies on the compatible block than on the incompatible block were assumed to indicate stronger negative than positive associations with Moroccan names (see

Greenwald et al., 1998), which was interpreted as reflecting higher levels of implicit prejudice.

4.1.2 Results

Implicit prejudice

Incorrect trials and the first two trials of each block were discarded from the ST-IAT analysis. Latencies below 300 ms were set to 300 ms (0.04%). Latencies above 3000 ms were set to 3000 ms (0.23%). Analyses were performed on log-transformed latencies, but untransformed mean latencies are reported (in milliseconds). An ST-IAT score was calculated for each participant by subtracting the average response latency in the compatible block ($M = 694, SD = 110$) from the average response latency in the incompatible block ($M = 737, SD = 125$), excluding one participant whose ST-IAT score was higher than 3 *SD* above the mean. A higher ST-IAT score therefore indicates relatively stronger negative than positive associations with Moroccan names (i.e., being higher in implicit prejudice). On average participants indeed had stronger negative than positive associations with the category of Moroccans, $t(92) = 5.04, p < .01, d = 0.37$.

Categorization task

For each participant the percentage faces categorized as Moroccan was calculated separately for each face set. A GLM analysis with face set (Moroccan-looking vs. criminal Moroccan-looking) as within-subjects factor and implicit prejudice as continuous factor revealed a main effect of face set,

$F(1, 90) = 64.29, p < .01, \eta^2_{\text{partial}} = .42$, such that participants categorized criminal Moroccan-looking faces ($M = 52.93, SD = 19.41$) more often as Moroccan than Moroccan-looking faces ($M = 45.64, SD = 18.87$). Additionally a main effect of implicit prejudice, $F(1, 90) = 9.70, p < .01, \eta^2_{\text{partial}} = .10$, and a Face Set x Implicit Prejudice interaction was found, $F(1, 90) = 5.23, p = .03, \eta^2_{\text{partial}} = .06$. In Figure 4.5a on page 82 the average percentage of criminal Moroccan-looking and Moroccan-looking faces categorized as Moroccan is plotted at -1*SD* and +1*SD* levels of implicit prejudice. As can be seen in Figure 4.5a, all participants categorized criminal Moroccan-looking faces more often as Moroccan than Moroccan-looking faces, but more prejudiced participants did so to a greater extent.

4.1.3 Discussion

As predicted participants allocated a larger percentage of the criminal-looking faces to the Moroccan category than of the non-manipulated faces. Moreover, more prejudiced participants showed this effect to a greater extent. This bias of highly prejudiced people towards over-allocation of criminal-looking faces to the Moroccan category supports our theoretical proposition: Highly prejudiced individuals more strongly endorse the stereotype that Moroccans are criminal and therefore expect Moroccan faces to look more criminal (as has been shown by Dotsch et al., 2008). Because these expectations affect normative fit, all people are more likely to categorize more criminal-looking faces as Moroccan, but highly prejudiced people do so to a greater extent. This hypothesized effect of prejudice on categorization has been clearly demonstrated in the present study.

However, it may be argued that the obtained effects are primarily based on valence. Ruys et al. (2008) proposed an evaluative fit hypothesis such that social categorization should be facilitated for exemplars that are evaluatively congruent with the target category. For example, Ruys et al. showed that stimulus persons' attractiveness (a positive feature) enhanced the speed of categorizing brides (a positive category) but not prostitutes (a negative category). In a similar vein, prejudiced participants in our study might have been more prone to allocate any negative-looking face to a negatively evaluated category, regardless of whether the negative features were stereotypical or not. In that case, normative fit would coincide with evaluative fit. The next study was designed to refute the idea that valence in itself (i.e, any negative trait) suffices to facilitate category allocation to the Moroccan category.

4.2 Study 2

We propose that evaluative fit plays a substantial role in category selection, on the condition that it originates from stereotype-relevant dimensions. In the Ruys et al. (2008) study the categorization of the negative category prostitutes was not influenced by the valence originating from attractiveness, but the positive category brides was. As Ruys and colleagues already alluded to in their discussion of these results, the lack of facilitation for unattractive prostitutes might be caused by the relative stereotype-irrelevance of attractiveness for that category. Stereotypes contain traits that people believe are predictive of category membership (Le Pelley et al., 2010). Walking on the street in daily life, the category prostitute will not be frequently activated for (un)attractive-looking people (we hope). Therefore, attractiveness is not predictive of being a prostitute and will not be part of the prostitute stereotype. Brides on the other hand often look their best on their wedding day, causing a far stronger

association with attractiveness. Because attractiveness is stereotype-relevant for brides, participants in the Ruys et al. study were faster in allocating attractive women to the bride category than unattractive women. Because attractiveness was stereotype-irrelevant for prostitutes, participants in the same study were equally fast in allocating attractive and unattractive women to the prostitute category, despite evaluative congruence of unattractive women and prostitutes.

In line with this argument, we expect only facial features that are associated with stereotype-relevant negative traits to enhance over-allocation to the categorization of Moroccan faces. To test this, we conducted a second study involving the Moroccan category and two negative traits: criminal and stupid. Criminal is highly stereotype relevant: When asked to list the content of the cultural stereotypes of Moroccans, Dutch participants most frequently listed the trait criminal (70%; Gordijn et al., 2001). The negative trait stupid was not mentioned in the list of traits named by more than 20% of participants and was therefore concluded not to be part of the shared cultural stereotype of Moroccans. In Study 2, we expected only facial features related to the stereotype-relevant trait criminal to elicit over-allocation of faces to the Moroccan category, but not facial features related to the stereotype-irrelevant trait stupid, despite both traits being negative and evaluatively congruent with the Moroccan category.

4.2.1 Method

Participants

Twenty male and 113 female Dutch-speaking students of the Radboud University Nijmegen participated in this study ($M_{\text{age}} = 21.67, SD = 4.73$). In return, participants received course credit or €4.

Overview and design

Participants were asked to categorize faces as either Moroccan or non Moroccan. These faces were Moroccan-looking, criminal Moroccan-looking, and stupid Moroccan-looking. Afterwards, implicit prejudice was measured. The experiment used a mixed model design with face set (Moroccan-looking vs. criminal Moroccan-looking vs. stupid Moroccan-looking) as within-subjects variable and implicit prejudice as continuous between-subjects variable. The dependent variable was the percentage of faces categorized as Moroccan.

Materials

Stupid and criminal features In order to manipulate stupid-looking facial features, a pilot study was run to construct noisy images of what participants thought a typical stupid face looks like. In this pilot, 3 male and 28 female students of the Radboud University Nijmegen ($M_{\text{age}} = 21.03, SD = 3.33$) completed the same forced-choice reverse-correlation image classification task as in the pilot of Study 1, but were instructed to select the most stupid-looking face. Trials on which participants responded faster than 300 ms were excluded. Six

participants were not included in the resulting averaged classification image, because they responded faster than 300 ms on more than 10% the 770 trials. Figure 4.3 shows the resulting averaged classification image, which represents what participants thought a typical stupid face looked like.

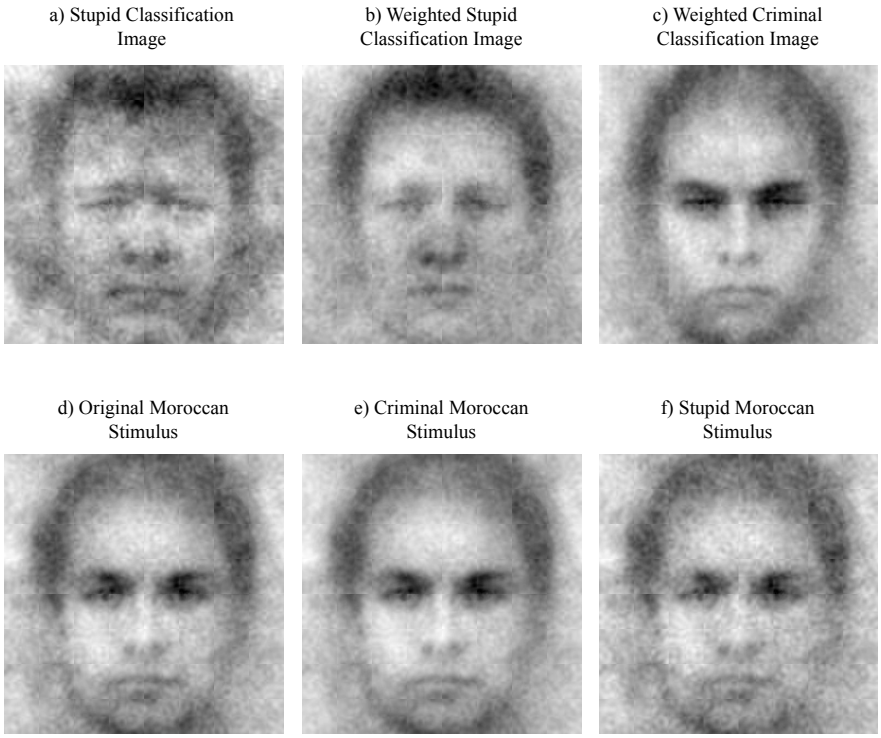

Figure 4.3: Stupid classification image (a), weighted stupid classification image (b), weighted criminal classification image (c), and example stimuli used in Study 2 (d, e, f)

To validate the stupid classification image, 26 male and 81 female students of the Radboud University Nijmegen ($M_{\text{age}} = 22.60, SD = 4.34$) rated all separate

30 criminal and 25 stupid classification images on criminality and stupidity on a 9-point scale ranging from 1 (*not criminal/not stupid*) to 9 (*very criminal/very stupid*). Ratings were averaged across participants for each image. The criminal faces were rated as more criminal ($M = 5.23, SD = 0.57$) than the stupid faces ($M = 4.94, SD = 0.39$), $t(53) = 2.17, p = .03, d = 0.59$. The stupid faces were rated as more stupid ($M = 5.02, SD = 0.31$) than the criminal faces ($M = 4.55, SD = 0.39$), $t(53) = 4.85, p < .01, d = 1.33$. However, the criminality and stupidity ratings of the stupid faces seemed to be correlated, $r(25) = .36, p = .08$. To filter out any characteristics perceived as criminal from the stupid face and vice versa, we orthogonalized the criminality and stupidity ratings and used the resulting values as weights to calculate weighted criminal and stupid classification images (see Figure 4.3b and 3c on page 65). As a result, the parameters underlying these classification images were uncorrelated, $r(4092) = -.04, n.s.$

To validate the weighted criminal and stupid classification images, 22 male and 109 female students of the Radboud University Nijmegen ($M_{\text{age}} = 21.40, SD = 3.97$) rated the two weighted classification images embedded in a sequence of six other noisy filler images on criminality and stupidity using the same scale as the previous rating study. The orthogonalization procedure seemed to have improved the extent to which the classification images conveyed the targeted traits: the weighted criminal face was rated as more criminal ($M = 7.76, SD = 1.35$) than the weighted stupid face ($M = 5.40, SD = 1.74$), $t(130) = 14.54, p < .01, d = 1.52$. The weighted stupid face was rated as more stupid ($M = 6.05, SD = 1.64$) than the weighted criminal face ($M = 5.48, SD = 2.06$), $t(130) = 3.25, p < .01, d = 0.31$.

Stimuli Three sets of stimuli were used in the experiment: 35 Moroccan-looking, 35 criminal Moroccan-looking, and 35 stupid Moroccan-looking faces.

The Moroccan-looking faces were the same as in Study 1 (see Figure 4.3d on page 65 for an example). The criminal and stupid Moroccan-looking faces were generated by respectively superimposing the weighted criminal and the weighted stupid classification image on each individual Moroccan-looking face (see Figure 4.3e and 4.3f). Again, the differences between the three classes of stimuli were very subtle.

Procedure

The procedure of Study 2 was the same as the procedure of Study 1, except for where indicated below.

Categorization task Participants this time categorized one by one and in random order 35 Moroccan-looking, 35 criminal Moroccan-looking, and 35 stupid Moroccan-looking faces into a Moroccan and a non-Moroccan category. There were two blocks and each face was presented once in each block, resulting in a total of 210 trials.

Implicit prejudice As prejudice measure, participants then moved on to the same ST-IAT as used in Study 1. However, due to time constraints, the ST-IAT this time consisted of only half the number of trials. Block order was kept constant: the compatible block always preceded the incompatible block.

4.2.2 Results

Implicit prejudice

Incorrect trials and the first two trials of each block were discarded from the ST-IAT analysis. Latencies below 300 ms were set to 300 ms (0.06%). Latencies above 3000 ms were set to 3000 ms (0.10%). Analyses were performed on log-transformed latencies, but untransformed mean latencies are reported (in milliseconds). A ST-IAT score was calculated by subtracting the average response latency in the compatible block ($M = 592, SD = 107$) from the average response latency in the incompatible block ($M = 627, SD = 122$), excluding one participant whose ST-IAT score was lower than 3*SD* below the mean. A higher ST-IAT score therefore indicates relatively stronger negative than positive associations with Moroccan names (i.e., being higher in implicit prejudice). As in Study 1, participants on average had stronger negative than positive associations with the category of Moroccans, $t(131) = 5.81, p < .01, d = 0.31$.

Categorization task

For each participant the percentage of faces categorized as Moroccan was calculated separately for each face set. A GLM analysis with face set (criminal Moroccan-looking vs. stupid Moroccan-looking vs. Moroccan-looking) as a three-level within-subjects factor and implicit prejudice as continuous factor revealed both a main effect of face set, $F(2, 129) = 60.33, p < .01, \eta^2_{\text{partial}} = .48$, and a Face Set x Implicit Prejudice interaction, $F(2, 129) = 3.74, p = .03, \eta^2_{\text{partial}} = .06$. Figure 4.5b on page 82 plots the results at -1 *SD* and +1 *SD* levels of implicit prejudice. Planned Helmert contrasts revealed that the implicit prejudice effect for criminal Moroccan-looking faces differed from the implicit pre-

judice effect for Moroccan-looking and stupid Moroccan-looking faces combined, $F(1, 130) = 4.21, p = .03, \eta^2_{\text{partial}} = .03$. The implicit prejudice effect did not differ between the latter two levels, $F(1, 130) = 0.06, n.s$. These effects were further explored in two separate GLM analyses with two-level face set factors: one with the levels Moroccan-looking vs. criminal Moroccan-looking faces and one with the levels Moroccan-looking vs. stupid Moroccan-looking.

Criminal faces A GLM analysis with face set (Moroccan-looking vs. criminal Moroccan-looking) as two-level within-subjects factor and implicit prejudice as continuous factor replicated the main effect of face set found in Study 1, $F(1, 130) = 51.92, p < .01, \eta^2_{\text{partial}} = .29$, such that participants categorized criminal Moroccan-looking faces ($M = 59.61, SD = 17.27$) more often as Moroccan than Moroccan-looking faces ($M = 52.22, SD = 17.39$). Additionally the Face Set x Implicit Prejudice interaction was replicated, $F(1, 130) = 6.55, p = .01, \eta^2_{\text{partial}} = .05$, in the expected direction (see Figure 4.5b on page 82): more prejudiced participants over-included more criminal Moroccan-looking faces in the Moroccan category.

Stupid faces A GLM analysis with face set (Moroccan-looking vs. stupid Moroccan-looking) as two-level within-subjects factor and implicit prejudice as continuous factor only revealed a main effect of face set, $F(1, 130) = 113.03$, $p < .01, \eta^2_{\text{partial}} = .47$, such that participants categorized stupid Moroccan-looking faces ($M = 41.17, SD = 22.02$) less often as Moroccan than Moroccan-looking faces ($M = 52.22, SD = 17.39$). There was no interaction with prejudice.

4.2.3 Discussion

As predicted, participants over-allocated criminal faces to the Moroccan category. Additionally, participants under-allocated stupid faces to the Moroccan category. These results go directly against the evaluative fit hypothesis. Categorization in Study 2 was affected by more than valence alone. Both criminality and stupidity are negative traits and therefore congruent with the negative valence of the Moroccan category. However only superimposing features associated with the stereotype-relevant trait criminal on Moroccan faces elicited over-allocation. Contrarily, superimposing features associated with the stereotype-irrelevant trait stupid evoked under-allocation. Thus, valence contributed to fit only to the extent that it derives from stereotype-relevant traits.

Moreover, more prejudiced participants allocated more criminal – but not stupid – Moroccan-looking faces to the Moroccan category. This suggests that the influence of implicit prejudice on category allocation is also restricted to faces with features associated with stereotype-relevant traits.

4.3 Study 3

In both Study 1 and 2 the trait criminal enhanced allocation of faces to the Moroccan category. Study 2 demonstrated that not any negative trait enhances over-allocation to a stigmatized category, but only negative stereotype-relevant traits. In Study 3 we aimed to test whether the same negative trait that elicits over-allocation to one stigmatized category, does not elicit it for another stigmatized category. Moreover, Study 3 included a positive stereotype-relevant trait, which, if it still causes over-allocation of faces conveying that trait for a

stigmatized category, would be even stronger evidence against the evaluative fit hypothesis, and in favor of a normative fit explanation.

Study 3 used the category of homosexual men, which can be perceived from faces (Rule & Ambady, 2008a; Rule, Ambady, Adams Jr, & Macrae, 2008), and two traits: the trait criminal – which is stereotype-irrelevant for homosexual men – and the stereotype-relevant trait feminine. Feminine is the most stereotypic personality trait ascribed to homosexual men (Madon, 1997). We expected participants to over-allocate feminine homosexual-looking faces to the homosexual category, but not criminal homosexual-looking faces. Because Study 2 showed that the influence of prejudice is restricted to stereotype-relevant traits, we also expected that if any effect of prejudice on category allocation would emerge, it would be obtained only for faces with features associated with the stereotype-relevant trait, i.e. feminine. We assumed feminine to be a positive trait, although the evidence for this is mixed. In many – but not all – cases, feminine or stereotypically female traits are judged to be more positive than masculine traits (e.g., Der-Karabetian & Smith, 1977; Silvern & Ryan, 1983, but see Eagly & Mladinic, 1994). Within the context of faces, a meta-analysis on attractiveness (Rhodes, 2006) showed that feminine female faces are clearly more attractive. More importantly for the current study, there is evidence for a preference for feminized male faces (Penton-Voak, Jacobson, & Trivers, 2004; Perrett et al., 1998; Rhodes, Hickford, & Jeffery, 2000), primarily in the case of using female facial averages to transform male facial images (Rennels, Bronstad, & Langlois, 2008; Rhodes, 2006). This is similar to the procedure in this study: superimposing a feminine classification image onto a male base face. Because femininity, in the context of the current stimuli, is a positive trait, we expected participants who are negatively prejudiced towards the homosexual category to allocate less of the feminine faces to the homosexual category. We expected prejudice towards homosexuals to be unrelated to the allocation of faces with features associated with the stereotype-irrelevant trait criminal.

4.3.1 Method

Participants

Three male and 42 female heterosexual Dutch-speaking students of the Radboud University Nijmegen participated in this study ($M_{\text{age}} = 21.76$, $SD = 2.52$). In return, participants received course credit or €4.

Overview and design

Participants were asked to categorize faces as either homosexual or non-homosexual. These faces were homosexual-looking, criminal homosexual-looking, and feminine homosexual-looking. Afterwards, implicit prejudice was measured. The experiment used a mixed model design with face set (homosexual-looking vs. criminal homosexual-looking vs. feminine homosexual-looking) as within-subjects variable and implicit prejudice towards homosexuality as continuous between-subjects variable. The dependent variable was the percentage of faces categorized as homosexual.

Materials

Homosexual and feminine features In order to create homosexual-looking faces and to manipulate feminine-looking facial features, a pilot study was run to construct noisy images of what people thought typical homosexual and feminine faces look like. In the pilot, 13 male and 54 female students of the Radboud University Nijmegen ($M_{\text{age}} = 22.19, SD = 5.25$) completed the same forced-choice reverse-correlation image classification task as in the previous

pilots. Thirty-five participants were instructed to select the most homosexual-looking face, whereas 32 other participants were instructed to select the most feminine-looking face. Trials on which participants responded faster than 300 ms were removed. Figure 4.4 shows the resulting averaged classification images, which represent what participants thought typical homosexual and feminine faces looked like.

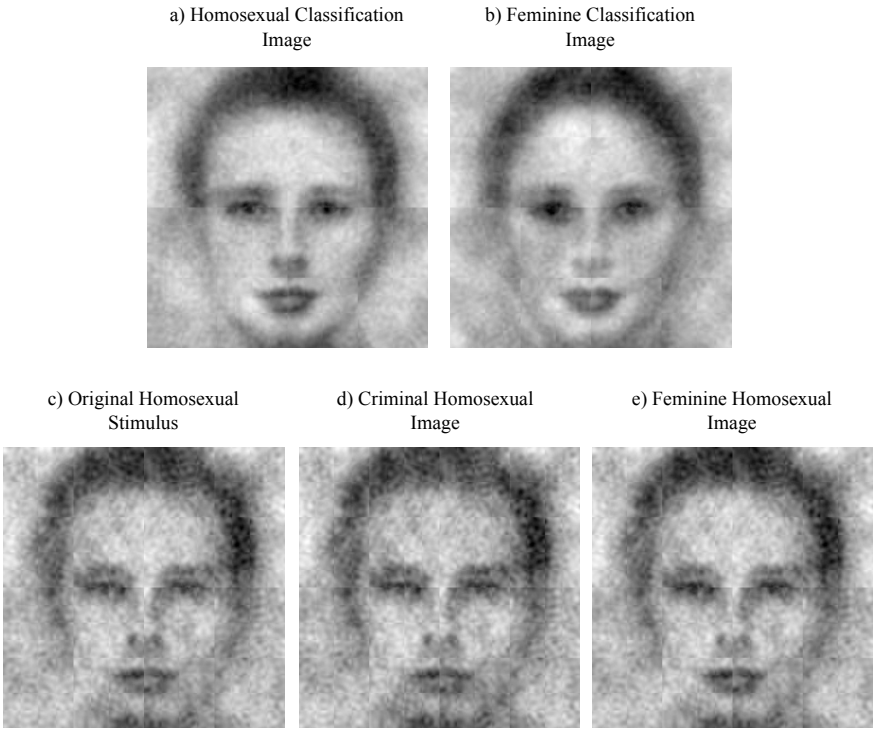

Figure 4.4: Homosexual classification image (a), feminine classification image (b), and example stimuli used in Study 3 (c, d, e)

Stimuli Three sets of stimuli were used in the experiment: 35 homosexual-looking, 35 criminal homosexual-looking, and 35 feminine homosexual-looking faces. The criminal classification image generated in the pilot of Study 1 was used for the criminal homosexual-looking faces. The faces were generated in a similar way as those in Study 1 (see Figure 4.4 on page 73 for example stimuli). Again, the resulting differences between the three classes of stimuli were very subtle.

Procedure

The procedure of Study 3 was the same as the procedure of Study 2, except for where indicated below.

Categorization task Participants categorized one by one and in random order 35 homosexual-looking, 35 criminal homosexual-looking, and 35 feminine homosexual-looking faces into a homosexual and a non-homosexual category. There were two blocks and each face was presented once in each block, resulting in a total of 210 trials.

Implicit prejudice As prejudice measure, participants then moved on to a homosexual-heterosexual IAT (Banse, Seise, & Zerbes, 2001; Steffens & Buchner, 2003). This task measured indirectly how strongly participants associated male homosexual couples (e.g., John + Richard) and heterosexual couples (e.g., Matthew + Julia), with positive and negative words. The IAT consisted of a compatible, incompatible, and three practice blocks, which were always carried out in the following order: In the first practice block participants classified 10 positive words with the left key and 10 negative words with the right. In

the compatible practice block participants classified 10 heterosexual couples with the left key and 10 homosexual couples with the right. In the following critical compatible block participants classified 10 positive words and 10 heterosexual couples with the left key and 10 negative words and 10 homosexual couples with the right key. The incompatible practice and critical block followed, with the functions of the keys with respect to the couples switched. Shorter response latencies on the compatible block than on the incompatible block were assumed to indicate stronger negative than positive associations with homosexual couples compared to heterosexual couples, which was interpreted as reflecting higher levels of implicit prejudice.

4.3.2 Results

Implicit prejudice

Incorrect trials and the first two trials of each block were discarded from the IAT analysis. Latencies below 300 ms were set to 300 ms (0%). Latencies above 3000 ms were set to 3000 ms (0.32%). Analyses were performed on log-transformed latencies, but untransformed mean latencies are reported (in milliseconds). An IAT score was calculated by subtracting the average response latency in the compatible block ($M = 710, SD = 122$) from the average response latency in the incompatible block ($M = 830, SD = 214$). A higher IAT score therefore indicates relatively stronger negative than positive associations with homosexual couples compared to heterosexual couples (i.e., being higher in implicit prejudice). Participants on average had stronger negative than positive associations with the homosexual category relative to the heterosexual category, $t(45) = 5.65 p < .01, d = 0.69$, i.e., the category of homosexuals is a stigmatized category on the implicit level for this sample.

Categorization task

For each participant the percentage faces categorized as homosexual was calculated separately for each face set. A GLM analysis with face set (feminine homosexual-looking vs. criminal homosexual-looking vs. homosexual-looking) as three-level within-subjects factor and implicit prejudice as continuous factor revealed a main effect of face set, $F(2, 42) = 107.31, p < .01, \eta^2_{\text{partial}} = .84$, and a marginally significant Face Set x Implicit Prejudice interaction, $F(2, 42) = 2.67, p = .08, \eta^2_{\text{partial}} = .11$. Figure 4.5c on page 82 plots the results at -1 *SD* and +1 *SD* levels of implicit prejudice. Planned Helmert contrasts revealed that the implicit prejudice effect for feminine homosexual-looking faces differed from the implicit prejudice effect for homosexual-looking and criminal homosexual-looking faces combined, $F(1, 43) = 5.17, p = .03, \eta^2_{\text{partial}} = .11$. The implicit prejudice effect did not differ between the latter two levels of face set, $F(1, 43) = 1.69, n.s.$ These effects were further explored in two separate GLM analyses with two-level face set factors: one with the levels homosexual-looking vs. feminine homosexual-looking faces and one with the levels homosexual-looking vs. criminal homosexual-looking.

Feminine faces A GLM analysis with face set (homosexual-looking vs. feminine homosexual-looking) as within-subjects factor and implicit prejudice as continuous factor revealed the expected main effect of face set, $F(1, 43) = 101.26, p < .01, \eta^2_{\text{partial}} = .70$, such that participants categorized feminine homosexual-looking faces ($M = 49.75, SD = 16.29$) more often as homosexual than homosexual-looking faces ($M = 32.48, SD = 13.77$). Additionally, a predicted Face Set x Implicit Prejudice interaction was marginally significant, $F(1, 43) = 3.86, p = .06, \eta^2_{\text{partial}} = .08$: more prejudiced participants, over-allocated feminine homosexual-looking faces to a lesser extent to the homosexual category (see Figure 4.5c on page 82).

Criminal faces A GLM analysis with face set (homosexual-looking vs. criminal homosexual-looking) as within-subjects factor and implicit prejudice as continuous factor only revealed a main effect of face set, $F(1, 43) = 122.32, p < .01, \eta^2_{\text{partial}} = .74$, such that participants categorized criminal homosexual-looking faces ($M = 19.05, SD = 10.93$) less often as homosexual than homosexual-looking faces ($M = 32.48, SD = 13.77$). As predicted, there was no interaction with prejudice.

4.3.3 Discussion

As expected, participants over-allocated feminine faces and under-allocated criminal faces to the homosexual category. These results demonstrate that the categorization of faces from different stigmatized categories relies on different trait dimensions conveyed by the faces: whereas the task of categorizing faces as Moroccan in Studies 1 and 2 elicited over-allocation of criminal faces, the task of categorizing faces as homosexual elicited under-allocation of criminal faces. Thus, faces with features associated with negative traits are not necessarily more likely to be allocated to a negatively evaluated category. This is a strong argument in favor of the stereotype-specificity of normative fit, and against the pure evaluative fit position. Note in this respect that the criminal features superimposed on the homosexual faces were, in fact, the very same features that were superimposed on Moroccan faces in Study 1.

Additionally, implicit prejudice towards homosexuals moderated the allocation of only the feminine faces, such that more prejudiced participants showed less over-allocation of feminine homosexual-looking faces. We have to interpret this effect with caution, because the interaction effect was only marginally significant. However, the finding sheds more light on the way normative fit

affects categorization. Faces that convey stereotype-relevant evaluative congruence have better fit for highly prejudiced individuals. In the case of the homosexual category the negativity associated with criminality derives from a trait that is not stereotype-relevant and therefore does not provide better fit for prejudiced individuals. Femininity, on the other hand, is stereotype-relevant, but positively valenced. Because prejudiced individuals evaluate the category as more negative, faces conveying less of the stereotype-relevant positive trait attractiveness have better fit.

4.4 General Discussion

In three studies we have shown that social categorization is biased at the level of category allocation. In Study 1 participants over-allocated faces conveying the stereotype-relevant negative trait criminal to the Moroccan category, especially if they were prejudiced. In the two subsequent studies we have shown a double dissociation: In Study 2 the stereotype-irrelevant negative trait stupid did not elicit over-allocation, but the stereotype-relevant negative trait criminal did. In Study 3, using the stigmatized category homosexual, the previously used negative trait criminal did not elicit over-allocation, but the stereotype-relevant positive trait femininity did. These results demonstrate that normative fit is higher for faces conveying stereotype-relevant traits regardless of valence. Moreover, individual differences in implicit prejudice predicted the extent to which stereotype-relevant traits caused over-allocation: whereas more prejudiced people showed greater over-allocation of faces conveying negative stereotype-relevant traits, they showed less over-allocation of faces conveying positive stereotype-relevant traits. These results strongly support our normative fit hypothesis: there is better fit for faces with features that are evaluatively congruent on the condition that those features are associated with stereotype-relevant traits, especially for prejudiced individuals.

Our results are especially striking considering the very subtle manipulation of criminal, stupid, and feminine features. Looking at the example stimuli in Figure 4.2 on page 59, it is difficult to see any difference between the original Moroccan and the criminal-looking Moroccan images. Nonetheless, our subtle manipulation, which was based on pilot-participants' criminality judgments, is strong enough to elicit a difference in category allocation between highly and less prejudiced participants.

Another strength of the current work is the use of reverse correlation methods to construct Moroccan, homosexual, criminal, stupid, and feminine-looking faces. At no point in the current research were any subjective notions of the researchers about what, for example, a criminal face looks like able to influence the results. Bottom-up data-driven research methodologies are gaining traction in research on higher-level cognition (e.g., Dotsch et al., 2008; Langner, Becker, & Rinck, 2009; Oosterhof & Todorov, 2008) and have as primary advantage that we are testing what is in the participants' mind, not what is in the researchers' mind (see Mangini & Biederman, 2004). It is not what the researchers believe a criminal face looks like that biases social categorization, but what participants believe a criminal face looks like.

Our work utilized a paradigm in which participants were forced to decide whether or not to allocate faces to one specific category. However, social psychologists have theorized that categories are activated automatically on perception of category members (Allport, 1954; Devine, 1989; Dovidio, Evans, & Tyler, 1986, for an overview, see Macrae & Bodenhausen, 2000). In daily life, people are mostly not on the lookout for members of one specific category, as in our task. The cognitive system has ample choice of social categories to activate (e.g., Moroccan, homosexual, Dutch, Black, Mexican, female, social psychologist). Future research should address the extent to which the effects obtained generalize to spontaneous categorization given multiple categories to allocate to.

Our results are in line with findings within the emotion perception literature that racially ambiguous faces with an angry expression are more likely to be categorized as Black by more prejudiced people (Hugenberg & Bodenhausen, 2004; Hutchings & Haddock, 2008, also see Bijlstra et al., 2010). Based on the former, Hugenberg and Sacco (2008, p. 1057) suggested that

> "when attributes of a social target are consistent with the attributes of a particular social category (e.g., angry expressions with the Black stereotype), this can strongly influence category selection".

Whereas the previous studies focused on emotional facial expressions indirectly derived from stereotypical traits (from the stereotype of Blacks as aggressive to angry facial expressions), the present study shows that facial features representing the actual stereotypical traits have a direct influence on category allocation. Moreover, the current work extends previous work by demonstrating that category selection might sometimes be biased by prejudice in the opposite direction: for the homosexual category more prejudiced individuals are less prone to over-allocate faces conveying the positive stereotype-relevant trait feminine.

The current work further advances knowledge about the relationship between prejudice and stereotypes. As noted by Wittenbrink et al. (1997), the definition of prejudice as an individual's negative attitude towards an out-group is widely accepted (Ashmore, 1970; Dovidio & Gaertner, 1986; Stroebe & Insko, 1989). A stereotype on the other hand reflects culturally shared beliefs (Devine, 1989) or personal beliefs about an out-group (Krueger, 1996). Amodio and Devine (2006) provided evidence for the independence of stereotyping and prejudice. On the other hand, Wittenbrink et al. concluded that on the implicit level stereotyping and prejudice are intertwined (p. 271). We go beyond both conclusions by demonstrating in what manner stereotyping and preju-

dice might be intertwined: the biased allocation by prejudiced people thrives on stereotype-relevant valence.

The demonstrated biases in category allocation might function both as a stereotype-maintenance and a prejudice-maintenance device (as has been alluded to by (Dotsch et al., 2008, p. 980). The reported experiments showed a general tendency to bias categorization in the direction of a culturally shared stereotype, thereby supporting the propagation of a cultural meme (Dawkins, 1976): As feminine-looking homosexuals are more likely to be categorized as homosexual, the likelihood that a counter-stereotypical (e.g., less feminine-looking) exemplar will be categorized as homosexual becomes smaller. Without counter-stereotypical category members, there is no need to change the stereotype. Moreover, the reported experiments showed a specific tendency of prejudice to bias categorization in the direction of evaluative congruency: as prejudiced individuals are more likely to categorize more negative and less positive exemplars into stigmatized categories, the likelihood that these individuals will categorize evaluatively incongruent exemplars into stigmatized categories becomes smaller. If prejudiced people never encounter any positive members of stigmatized groups (because they do not categorize them as such), there is no reason to change their evaluation of those groups. They remain prejudiced.

Social categorization (Allport, 1954) has proven to be an important concept in explaining prejudice (see Fiske, 2005). Ample research has shown the influences of prejudice on the consequences of categorization. The current research clearly demonstrates the effect of prejudice on the categorization process itself.

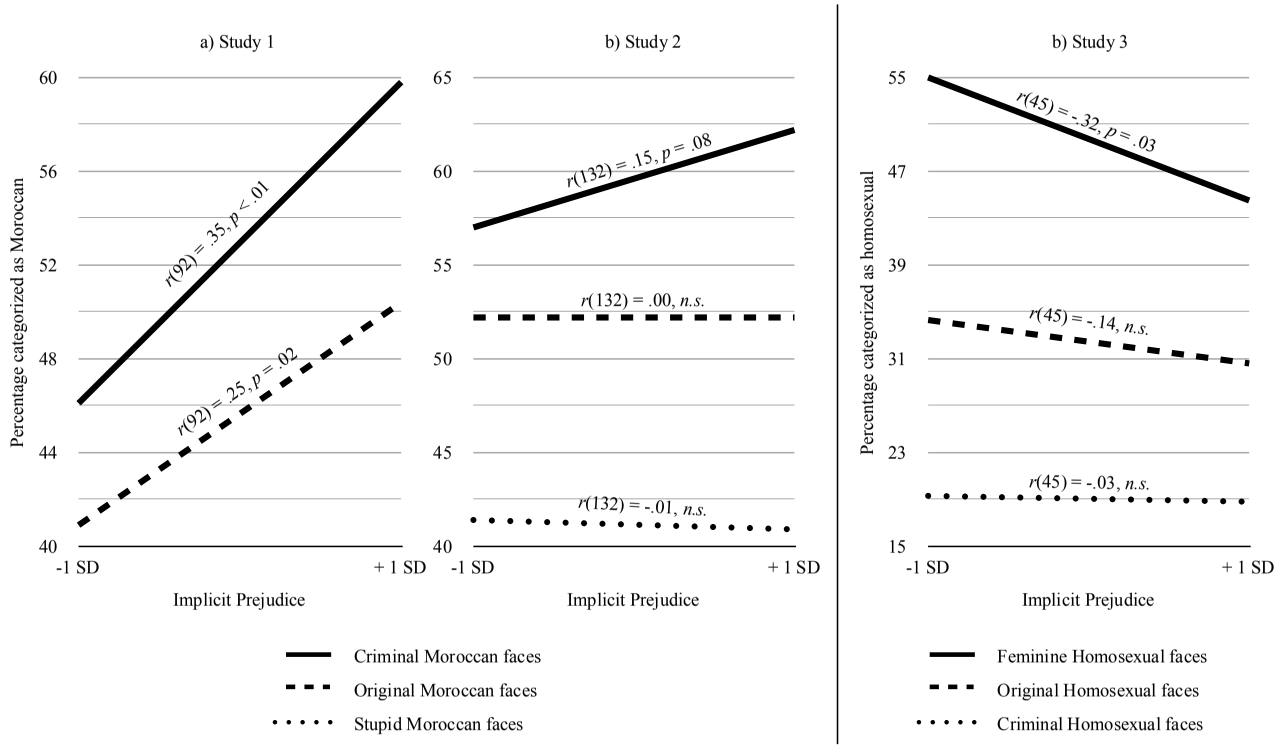

Figure 4.5: Average percentage of faces allocated to category by face set at -1 *SD* and +1 *SD* values of implicit prejudice in Study 1 (a), Study 2 (b), and Study 3 (c)

Chapter 5

General discussion

This dissertation aimed to test an extended model of social categorical perception, presented in the introduction. This extended model differs from mainstream theories of social categorization in that it emphasizes the category allocation phase in which perceptual input is mapped onto appropriate categories. According to the model, accessible categories are activated when they fit category specifications enclosed in the stereotype. The present dissertation focused on category specifications included in visual stereotype content. Because there are individual differences in stereotype content, we argued that people also differ in the visual stereotype content they have and therefore in their visual category specifications. As a consequence different people might activate different categories for the same perceived person under the same circumstances. According to the model, this variation in stereotype content is predicted by prejudice. Consequently, the model predicts that prejudice not only influences intergroup bias after categorization has taken place (as has been

shown repeatedly in the past, see Chapter 1), but also influences the categorization process itself, by biasing who gets placed into what category. Figure 5.1 on page 84 depicts the extended model. We tested this model across three empirical chapters. The studies in these chapters are discussed in relation to four major themes.

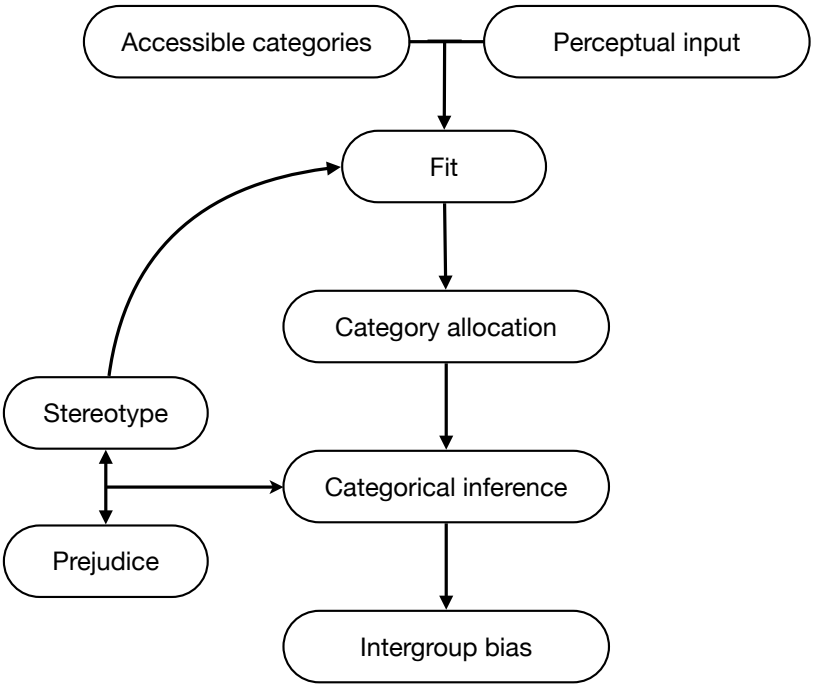

Figure 5.1: Extended model of categorical perception

5.1 Empirical support for the extended model

Two studies in Chapter 2 showed that Dutch participants who were more prejudiced towards Moroccans, expected typical Moroccans to look more criminal and less trustworthy. Within the framework of the extended model, this is evidence for the postulated link between prejudice and stereotype content (which includes expectations about typical group members). The link between prejudice and stereotype has been established previously within the domain of verbal stereotype content (Gordijn, Koomen, & Stapel, 2001; Wittenbrink, Judd, & Park, 1997). The studies in Chapter 2 established this link for the visual component of stereotypes. This is a necessary first step of testing the model, as we argued that visual stereotype content is mediating the effect of prejudice on the establishment of fit for a perceived person. Moreover, the studies demonstrated that reverse correlation methods have the capacity to lay bare visual stereotypes of stigmatized groups (we will come back to this point later).

From the results in Chapter 2 it is unclear whether the negative associations with Moroccans cause more criminal visual content or the other way around. Therefore, the study in Chapter 3 aimed to induce new associations about a novel group (group X) by asking participants to form an impression based on behavioral information about and facial appearance of group X members. We manipulated behavioral information, while keeping group members' facial appearance constant. The results showed that visual stereotype content was affected by the behavioral information manipulation. This effect was mediated by prejudice. Thus, Chapter 3 provided experimental evidence for the effect of prejudice on visual stereotype content. This finding is important, because it suggests that the bias in the expected facial appearance of typical group members is not purely a function of the actual facial appearance of encountered group members, but can be caused by group information unrelated to facial appearance.

In the study presented in Chapter 3 we used behavioral information of group exemplars as a source of bias. However, other sources are also possible, such as trait information about group members or the group as a whole. The reverse effect might also take place: because people read trait information from faces (e.g., Todorov et al., 2008), exemplar facial appearance might further bias attitudes and face representations, perhaps in interaction with behavioral information. Note that facial appearance was kept constant in Chapter 3.

Importantly, the bias in visual stereotype content was predicted by measures of explicit and implicit prejudice. This is, again, evidence for the link between prejudice and visual stereotype content. Whereas Chapter 2 provided evidence for this link for one specific real world stigmatized out-group (Moroccans), Chapter 3 showed that the effect is generalizable to a novel group, and can be induced after encountering just a few members of a group.

Having established that prejudice affects visual stereotype content, in Chapter 4 we tested the implications for social categorization. According to our extended model, prejudice affects fit between accessible categories and perceptual input through visual stereotype content. Faces that better fit the category specifications represented in the visual stereotype, have a higher probability to be categorized as such. Indeed, Study 4.1 demonstrated that faces with facial features associated with stereotypical traits (in this case criminal) are more likely to be categorized as Moroccan. This effect was stronger for participants high in prejudice. This is in line with the findings in Chapters 2 and 3 that highly prejudiced participants have a more stereotypical (criminal) representation of Moroccan faces. Because highly prejudiced participants have a more criminal visual stereotype, fit with the Moroccan category is enhanced for criminal-looking faces.

Taken together these chapters provide strong evidence for the effect of prejudice on visual stereotypes, and indirectly, for the effect of visual stereotypes on fit and categorization, thereby supporting our extended model. However, several important issues are raised by the model and the empirical chapters, which we will now turn to.

5.2 Valence or traits?

We argued that the visual stereotype contains the typical facial appearance of group members, comprising of facial features associated with specific stereotypical traits. However, the traits we employed in Chapters 2 and 3, *trustworthy* and *criminal*, clearly differed in valence: trustworthy is positive and criminal is negative. Because also prejudice, i.e., the negativity associated with a category (Dovidio & Gaertner, 2010; Fiske, 1998), is an evaluative construct, one could argue that we have merely provided evidence for an evaluative account of influences on the early category allocation phase. In such an account, prejudice is related to visual stereotype negativity, and as a consequence more negative looking faces will have enhanced fit for the category. We believe such an account to be too simplistic, both methodologically and theoretically.

First of all, it is important to note that the reverse correlation procedure can in principle produce visualizations of representations that differ on dimensions other than valence. For instance, Oosterhof and Todorov (2008) have used reverse correlation procedures to visualize people's representations of personality trait dimensions such as trustworthiness, dominance, and competence. Not all of these traits covary with valence. In fact, Oosterhof and Todorov used principal component analysis to demonstrate that, whereas trustworthiness converged highly with the first principal component (which seemed to code for

valence), dominance converged highly with a second component, orthogonal to valence. We replicated the visualization of these trait dimensions using the same reverse correlation procedure as used in this dissertation (Dotsch, Todorov, & Wigboldus, in prep). Thus, our procedure to assess visual stereotype content is in principle capable of producing differences on other dimensions than valence alone.

Moreover, theoretically it makes sense to differentiate stereotype content in terms of more than valence alone. For instance, Fiske, Cuddy, Glick, and Xu (2002), in the validation of their influential stereotype content model, convincingly demonstrated that stereotype content is at the least represented by two dimensions: warmth and competence. The first dimension, warmth, might be considered a highly evaluative dimension. Likewise, Oosterhof and Todorov (2008) demonstrated that a two-dimensional model of face evaluation represents about 80% of trait variance inferred from faces (as opposed to 63% for the first dimension alone, coding for valence). Thus, stereotype content is multi-dimensional (Fiske et al., 2002), and given that visual stereotype content includes facial representations, those representations too can contain multi-dimensional social information (Oosterhof & Todorov, 2008). Thus, there is no reason for visual stereotype content to be limited to a uni-dimensional concept such as valence. Although stereotype content might be described to some extent by valence, assessing stereotype content solely with valence does not do justice to a system of categories that goes beyond the good vs. bad dimension.

In short, our measure of visual stereotype content is not limited to valence and there is no theoretical reason to assume that visual stereotype content is limited to valence. However, the relation between prejudice and visual stereotype content may still be limited to valence. Because prejudice is an affective or evaluative construct (Dovidio & Gaertner, 2010; Fiske, 1998), it likely covaries primarily with valenced stereotype content. Importantly, prejudice is not

a perfect predictor of visual stereotype content. In Chapter 2 we showed that for the stigmatized category of Moroccans (or for a clearly negative or positive novel category in Chapter 3), prejudice is a predictor of the extent to which the visual stereotype contains facial features associated with negative stereotypical traits. However, as argued above, visual stereotype content might additionally contain information about multiple trait dimensions, some of which are orthogonal to valence. These dimensions, although part of visual stereotype content, are unrelated to prejudice.

Let us specify our view on the relation between prejudice and visual stereotype content in more detail. Prejudice predicts the extent to which the visual stereotype contains facial features associated with valence-laden traits. Additionally, the visual stereotype may contain facial features associated with unvalenced traits. The extent to which the visual stereotype contains the latter facial features is not predicted by prejudice. Moreover, which valenced (or unvalenced) traits are represented in the visual stereotype is not necessarily related to prejudice, but might for instance depend on the belief that certain specific traits are predictive of group membership (see Le Pelley et al., 2010). This view is consistent with Stroebe and Insko (1989) and Wittenbrink et al. (1997) within the domain of verbal stereotype content.

The studies in Chapter 4 illustrate this relationship. The trait criminal is stereotypical for Moroccans and therefore is represented in the visual stereotype (Studies 2.1 and 2.2). Because the trait criminal is negative, prejudice predicts the extent to which the trait is represented in the visual stereotype, and ultimately the likelihood that a criminal-looking face is categorized as Moroccan (Studies 4.1 and 4.2). Other negative traits, such as stupid, are not stereotypical for the category of Moroccans (e.g., Gordijn et al., 2001) and are therefore not represented in the visual stereotype. Consequently, prejudice is unrelated to the likelihood that a stupid-looking face is categorized as Moroccan (Study

4.2). Likewise, the trait feminine is stereotypical for the category of homosexuals and therefore represented in the visual stereotype. Because the trait is valenced, prejudice predicts the extent to which the trait is represented in the visual stereotype (Study 4.3). Because the trait is positive (see Chapter 4), prejudiced people are less likely to have a visual stereotype with feminine facial features (or have a visual stereotype with less feminine facial features), and therefore are less likely to categorize feminine-looking faces as homosexual than less prejudiced people. Another valenced trait, criminal, is not stereotypical for the category of homosexuals and is therefore not represented in the visual stereotype. Consequently, prejudice does not affect the likelihood that a criminal-looking face is categorized as homosexual (Study 4.3).

In short, we suggest in answer to the valence versus traits problem that higher prejudice entails stronger negative (and weaker positive) stereotypical features in visual mental representations of group members and that this relationship is constrained by descriptive stereotype content.

5.3 Visual stereotype or semantic stereotype?

We proposed, in line with Brewer (1988), that stereotypes have a visual component, which among other things represents typical group members' faces. To what extent did the present empirical work provide evidence for this proposition? One may prematurely conclude that the Moroccan classification images resulting from reverse correlation tasks (such as those presented in Chapter 2) represent an approximation of participants' visual stereotype content and that, therefore, this research proves the existence of a visual component of stereotypes. However, the fact that the task relied on pictures on a computer screen does not imply that participants actually based their responses on pictures in

their head (i.e., visual stereotype content). To claim that the reverse correlation task measured visual stereotype content is equivalent to claiming that participants matched their visual representation of typical Moroccan faces to the stimuli on the screen, *without activating the semantic stereotype content* (e.g., the trait criminal). However, an equally plausible explanation is that participants activated the semantic stereotype content about Moroccans (e.g., the trait criminal) and subsequently selected the most criminal-looking face, without matching any visual input to visual stereotype content.

Based on our data, we can not directly distinguish between the two processes. The best way to pit these two processes against each other is to prevent activation of semantic stereotype content during the reverse correlation task. One possibility would be to conduct a replication of the studies in Chapter 2, in which participants generate a Moroccan classification image, while the semantic activation of stereotypical traits is blocked. This could be achieved using modality-specific cognitive load (e.g., Rees, Frith, & Lavie, 2001). Participants would be asked to perform two tasks at the same time. One task is the reverse correlation described in Chapter 2. Every trial of the reverse correlation task would be preceded by a trial of the second task, a 2-back working memory load task (Jonides et al., 1997), designed to preoccupy either the verbal or visual modality during the reverse correlation trial. In 2-back trials participants would see a picture (in the visual block condition) or the name (in the verbal block condition) of an object. Their task would be to decide whether the current object is the same as the object that was presented two 2-back trials earlier. Modality-specific working memory thus would continuously be occupied by two objects, rendering it unavailable for use during reverse correlation trials.

If participants indeed rely on their visual stereotype when performing the reverse correlation task, blocking the verbal modality should not affect the rela-

tionship between prejudice and criminality of the Moroccan classification image. That is, participants who are more prejudiced towards Moroccans should generate a more criminal-looking classification image than less prejudiced participants. Blocking the visual modality, on the other hand, should lead to a weaker relationship between prejudice and criminality of the Moroccan classification image, because participants would not be able to rely on the match between the perceived faces and their visual stereotype of Moroccans. This pattern of results would support the hypothesis that stereotypes include a distinct visual component.

If stereotypes are indeed represented by a visual component, in addition to a semantic component, these components should be discernible on the neural level using brain imaging techniques. Importantly, it has been shown that visual and semantic brain areas can be identified in the brain. Visual processing is primarily supported by areas of the visual cortex (such as V1) and there is a well-identified network of face selective regions in the posterior cortex, consisting of the fusiform face area (FFA), the occipital face area (OFA), and posterior superior temporal sulcus (pSTS) (Haxby et al., 2000; Said, Dotsch, & Todorov, 2010). During both visual perception and visual imagery, these areas are active. Amodal semantic processing, on the other hand, is primarily supported by the anterior temporal lobes (ATL; Patterson, Nestor, & Rogers, 2007), a severely degenerated area in patients suffering from semantic dementia. Over time, these patients lose knowledge about properties of everyday objects in the context of otherwise well-preserved cognition (e.g., Warrington, 1975).

Based on the fact that different brain areas are involved in semantic versus visual processing, the following experiment could be conducted using fMRI to get more insight into what kinds of representations are being used by participants while performing the reverse correlation task. A reverse correlation task identical to the one in Chapter 2 would be followed by a semantic task

and a visual task. In the semantic task, participants would repeatedly choose from two traits the trait which according to them is most stereotypical for Moroccans. In the visual task, participants would repeatedly decide for two faces whether they belong to the same ethnicity. These faces would be drawn from a pool of faces with features associated with several ethnicities, half of which Moroccan.

The reverse correlation task is hypothesized to lay bare visual stereotype content, to the extent that it is related to prejudice. Whether this is the case, can be inferred from spontaneous brain activity while participants perform the reverse correlation task. The brain activity in the reverse correlation task could be decomposed into activity specific to semantic processing (based on the overlap with activity during the semantic task) and activity specific to visual processing (based on the overlap with activity during the visual task). If there is significant visual brain activity during the reverse correlation task, and this activity covaries with classification image criminality and prejudice, then there is strong neural evidence for a distinct visual component of stereotypes captured by the reverse correlation task.

In the introduction we made a theoretical case in favor of the existence of visual stereotypes, but we did not provide evidence supporting this contention. Instead, we assumed that stereotypes have visual components based on previous research, and have suggested a promising method to assess those components (i.e., the reverse correlation procedure). This is important groundwork, but experiments as those proposed above are required to specifically address the existence of visual stereotype components. Alternatively, we focused on the consequences of biases in visual stereotypes for social categorization due to individual differences in prejudice. Even if ultimately stereotypes turn out to be of purely semantic or verbal nature, our results are not invalidated. Prejudice still influences social categorization at the level of category allocation. But, the

underlying process would differ from the one currently suggested. However, as argued in the general introduction, we believe that visual stereotype content provides a more parsimonious model for the data in the present dissertation than purely semantic or verbal stereotype content.

5.4 Psychophysical reverse correlation methods

This dissertation introduces data driven reverse correlation methods to the field of stereotyping and prejudice. These methods hold much promise for social psychology, as they allow researchers to probe internal representations without imposing their own ideas on what those representations should look like (Gosselin & Schyns, 2003). In Chapter 4, the reverse correlation method was used to visualize criminal faces. Conventional experimental paradigms would have required us to manipulate the specific facial feature dimensions that we hypothesize to be relevant for the perception of criminality. As a result, we would risk manipulating dimensions that are irrelevant, committing us to a larger cycle of repeated experiments. Each time we would have to think of other potentially relevant dimensions and then test whether those affect perception of criminality until we finally find dimensions that do. Even then, we would never be sure whether the relevant dimensions that we end up with are the only ones that affect perception of criminality, if we do not bother to test more dimensions afterwards. Reverse correlation methods speed up scientific discovery by offering a data-driven method to identify relevant dimensions in just one experiment, while minimizing the influence of researchers on the end result (as stimuli are randomly generated).

Moreover, reverse correlation methods offer the possibility to quantify *spontaneous* processes. For instance, in the reverse correlation task in Chapter 2, we

asked participants to select the most Moroccan looking face. At no point before or during the reverse correlation task was criminality (or valence) mentioned to the participants. Unlike most contemporary indirect measures, participants are not forced by the paradigm to consider attribute dimensions (valence or traits) pertaining to the category of interest. If participants do not think about Moroccans in terms of criminality or negativity, they will be able to complete the reverse correlation task without activating those dimensions. Compare this to a Moroccan-Dutch IAT (e.g., Greenwald et al., 1998). In one block participants are asked to pair Moroccan names with negative words and Dutch names with positive names, and in the other they are asked to pair Dutch names with negative words and Moroccan names with positive words. The IAT clearly forces participants to adopt an evaluative mindset. Contrarily, in reverse correlation tasks participants are simply asked to select faces, a decision which can be based on any dimension a participant deems relevant.

We have successfully employed these methods in several other social psychological domains. For instance, Karremans, Dotsch, and Corneille (in prep) have used the reverse correlation method to demonstrate so-called beauty blindness. Romantically involved and single female participants were asked to memorize the face of a potential mate who had either an attractive or unattractive facial appearance. Subsequently, their mental representation of the alternative mate was visualized using a reverse correlation task (using the same task as in this dissertation). The resulting classification images were rated on attractiveness by independent participants. Classification images generated by romantically involved participants showed less differentiation on attractiveness for the attractive versus unattractive alternative mate than those generated by single participants.

Imhoff, Dotsch, Bianchi, Banse, and Wigboldus (in prep) used the reverse correlation technique to visualize ingroup projection. According to the Ingroup

Projection Model, people who belong to a group tend to generalize typical ingroup characteristics to the superordinate category (Mummendey & Wenzel, 1999). Imhoff and colleagues investigated whether the typical facial appearance of members of one's ingroup is generalized to the superordinate category. To test this hypothesis, participants in Germany and Portugal completed a reverse correlation task in which they repeatedly had to select the most European-looking face. German participants generated European classification images that according to Dutch raters looked more German and less Portuguese than the European classification images generated by Portuguese participants.

These are just two examples of the potential of reverse correlation as a social psychological research tool. Although the possibilities seem endless, several limiting factors of reverse correlation methods should be considered. First, as discussed above, it is unclear what the reverse correlation method is actually assessing. The resulting classification image could be an approximation of underlying mental representations, but could arguably also be a simple response strategy, unrelated to any mental representations. Even if the procedure does tap into underlying mental representations, resulting classification images will always be a function of those mental representations, the used base face, the specific outcome space spanned by the presented noise patterns, and error. Second, the resulting classification image does not reveal much about the psychological properties of the assessed mental representation, which could be a collection of exemplars, a feature list, a prototype, etc. Third, in order to generate high quality visualizations, participants need to respond to a large number of trials (the smallest number of trials being 300, used by Oosterhof & Todorov, 2008; the largest being 20.000, used by Gosselin, Bacon, & Mamassian, 2004). However, if the number of trials is too high, the quality of the visualizations deteriorates as well, due to participant demotivation. Fourth, because reverse correlation entails a linear analysis which does not take interactions into ac-

count, artifacts in the resulting classification images emerge. Mangini and Biederman (2004) illustrate this with the example of an eye-wink. If participants are asked to select any face that shows an eye-wink in a reverse correlation task, they will probably select any face with a closed left eye or a closed right eye. As a consequence, the resulting classification image, calculated as the average of all selected images, will show a half closed left eye and a half closed right eye. Needless to say, for all the reasons above, researchers should take care with interpreting classification images as direct visualizations of mental representations.

Future developments in reverse correlation techniques might make it possible to model interactions as well as nonlinear effects. This is difficult with the method used in this dissertation because the data of at most 770 responses were used to estimate the values of 4096 parameters. If interactions between parameters would need to be taken into account, over 8 million terms should have been estimated, clearly causing overfitting. One way to counter this is limiting the number of parameters to those that actually code for changes in the face (a lot of parameters in the task used in the present dissertation code for changes outside of the face). Oosterhof and Todorov (2008) solved this issue by using random changes in 50-dimensional face space, thereby limiting the number of variables to estimate to 50, and the number of first-order interaction terms to 1225. Although the number of trials used by Oosterhof and Todorov – 300 – is still too low for taking into account interactions, this development is promising. Moreover, the number of trials necessary for high quality classification images might be dramatically improved when researchers employ more efficient search algorithms, which base the presentation of following stimuli on responses to previous stimuli.

Another recent development is the multiple response alternative reverse correlation task (Dai & Micheyl, 2009), which allows for the visualization of mul-

tiple categories in the same task. For instance, within the context of the present dissertation, participants could be asked to complete a reverse correlation task with ten different ethnic categories instead of just Moroccan. This would enable us to identify whether the biasing effect of prejudice on the visual stereotype generalizes across other ethnic groups and whether there is a common factor causing this bias across groups within a single person.

In its current form, reverse correlation is already gaining momentum in social psychology. In the context of face perception, these methods have been applied successfully to perception of gender (Nestor & Tarr, 2008), identity (Mangini & Biederman, 2004), expressions of emotions (Langner et al., 2009; M. L. Smith, Cottrell, Gosselin, & Schyns, 2005), personality traits (Oosterhof & Todorov, 2008), and even cross-species categories (i.e., baboon faces; Martin-Malivel, Mangini, Fagot, & Biederman, 2006; for an in-depth discussion of reverse correlation paradigms, see Mangini & Biederman, 2004; Todorov, Dotsch, Wigboldus, & Said, under review).

5.5 Conclusion

The extended model of categorical perception proposed here (see Figure 5.1 on page 84) builds on a strong foundation of social psychological theory and experiments. The model follows ideas originally articulated by Bruner (1957), namely that fit with specifications of accessible categories predicts which category will be activated for a perceived stimulus. Our model integrates theories from object categorization and person perception by stating that these specifications are part of the stereotype content (hinted at by e.g., Blanz, 1999; van Knippenberg & Dijksterhuis, 2000; McGarty, 1999). Moreover, our model incorporates recent insights of knowledge representation (e.g., Barsalou, 1999)

by postulating that stereotype content consists not only of purely semantic or verbal information, but also of perceptual –and specifically visual– information. The model was tested within the domain of face perception, capitalizing on the strong need within social psychology to use more ecologically valid stimuli (Brewer, 1988; Carlston, 1994; Feldman, 1988; McArthur & Baron, 1983; McGarty, 2002; Zebrowitz, 1996, 2006; Zebrowitz & Montepare, 2008), especially in person perception research, in which the use of verbal category labels as stimuli has been the rule more than the exception.

Traditionally, social psychologists treated social categorization as a process that biases subsequent perception, cognition, and behavior. The extended model on the other hand entails the proposition that the categorization process itself is subject to biases, moderated by prejudice and constrained by visual stereotype content. The empirical work in the previous chapters supports the idea that social categorization is biased by prejudice and suggests a crucial role for visual stereotype content in this process. However, as discussed above, the work is not yet done. Specifically, the existence of distinct visual or perceptual stereotype content above and beyond verbal or semantic stereotype content is still tentative. In applying reverse correlation paradigms to stereotype research, the present work provides new tools to tackle the age old problem of conceptual knowledge representation (e.g., Fodor, 1998) in a social context.

After having discussed the theoretical and methodological implications and insights, one important question is left unasked: Why does the bias in social categorization occur? We propose that the visual stereotype and categorization biases identified in this dissertation fuel stereotype and prejudice maintenance. The reported experiments in Chapter 4 showed a general tendency to bias categorization in the direction of a culturally shared stereotype, thereby supporting the propagation of a cultural meme (Dawkins, 1976): As feminine-looking homosexuals are more likely to be categorized as homosexual, the likelihood

that a counter-stereotypical (e.g., less feminine-looking) exemplar will be categorized as homosexual becomes smaller. Without counter-stereotypical category members, there is no need for stereotype change to occur. And so the stereotype perpetuates.

Moreover, the experiments reported in Chapter 4 showed a specific tendency of prejudice to bias categorization in the direction of evaluative congruency: as prejudiced individuals are more likely to categorize more negative and less positive exemplars into stigmatized categories, the likelihood that these individuals will categorize evaluatively incongruent exemplars into stigmatized categories becomes smaller. If prejudiced people never encounter any positive members of stigmatized groups (because they do not categorize them as such), there is no reason to change their evaluation of those groups. They remain prejudiced.

In our extended model of categorical social perception, stereotypes play a fundamental role in perception, beyond filling in informational gaps about categorized individuals. That is, perceiving and understanding objects and people requires an informational construct which facilitates pattern matching (categorization) as well as pattern completion (stereotype activation). However, as argued in the introduction, in mainstream social psychology only one function – pattern completion – has become associated with the stereotype. The extended model is the first model to reinstate the stereotype's other function: pattern matching. Ultimately, Lippmann (1922) foresaw this function of stereotypes when he wrote that

> "for the most part, we do not see and then define; we define first and then see."

References

Aarts, H., Verplanken, B., & van Knippenberg, A. (1997). Habit and information use in travel mode choices. *Acta Psychologica*, *96*(1-2), 1–14.

Ahumada, A. J. (1996). Perceptual classification images from Vernier acuity masked by noise. *Perception*, *26*(18), 1831–1840.

Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, *2*(1).

Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, *49*(6), 1751–1756.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*(4), 249–277.

Ashmore, R. D. (1970). Prejudice: Causes and cures. In B. E. Collins (Ed.), *Social psychology: Social influence, attitude change, group processes and prejudice* (pp. 245–339). Reading, MA: Addison-Wesley.

Ashmore, R. D., Del Boca, F. K., & Titus, D. (1984). Types of women and men: Yours, mine, and ours. In *American psychological association meeting*,

toronto.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift f "ur experimentelle Psychologie*, *48*(2), 145–160.

Barsalou, L. W. (1983). Ad hoc categories. *Memory Cognition*(11), 211-227.

Barsalou, L. W. (1990). Access and inference in categorization. *Bulletin of the Psychonomic Society*, *28*(3), 268–271.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*(22), 577-609.

Beard, B. L., & Ahumada, A. J. (1998). Technique to extract relevant image features for visual tasks. In *Proceedings of spie* (Vol. 3299, p. 79).

Belmore, S., & Hubbard, M. (1987). The role of advance expectancies in person memory. *Journal of Personality and Social Psychology*, *53*(1), 61–70.

Bijlstra, G., Holland, R. W., Dotsch, R., & Wigboldus, D. H. J. (in prep). *The role of stereotype associations in emotion recognition.*

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. J. (2010). The Social Face of Emotion Recognition: Evaluations Versus Stereotypes. *Journal of Experimental Social Psychology*.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*(10), 674.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of personality and social psychology*, *83*(1), 5–25.

Blanz, M. (1999). Accessibility and fit as determinants of the salience of social categorizations. *European Journal of Social Psychology*, *29*(1), 43–74.

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European journal of social psychology*, *38*(6), 977–997.

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer Jr. (Ed.), *Stereotype activation and inhibition: Advances in*

social cognition (Vol. 11, pp. 1–52). Lawrence Erlbaum.

Brehmer, B., & Joyce, C. R. B. (1988). *Human judgment: The SJT view*. North Holland.

Brewer, M. B. (1988). A dual-process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition: A dual process model of impression formation* (pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brewer, M. B., & Feinstein, A. S. H. (1999). Dual processes in the cognitive representation of persons and social categories. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 255–270). New York: The Guilford Press.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152.

Bruner, J. S., Goodnow, J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Campbell, D. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, *3*(1), 14–25.

Carlston, D. E. (1994). Associated systems theory: a systematic approach to cognitive representations of persons. In T. K. Srull & R. S. Wyer Jr. (Eds.), *Advances in social cognition* (Vol. 7, p. 1-78). Hillsdale, NJ: Erlbaum.

Carver, C., & Garza, N. (1984). Schema-Guided Information Search in Stereotyping of the Elderly. *Journal of Applied Social Psychology*, *14*(1), 69–81.

Castelli, L., Carraro, L., Gawronski, B., & Gava, K. (2010). On the determinants of implicit evaluations: When the present weighs more than the past. *Journal of Experimental Social Psychology*, *46*(1), 186–191.

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The Perceptual Determinants of Person Construal: Reopening the Social-Cognitive Toolbox. *Journal of Personality and Social Psychology*, *88*(6), 885–894.

Coenders, M., Lubbers, M., Scheepers, P., & Verkuyten, M. (2008). More than two decades of changing ethnic attitudes in the Netherlands. *Journal of Social Issues*, *64*(2), 269–285.

Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When Memory Shifts Toward More Typical Category Exemplars: Accentuation Effects in the Recollection of Ethnically Ambiguous Faces. *Journal of Personality and Social Psychology*, *86*(2), 236–250.

Corneille, O., Hugenberg, K., & Potter, T. (2007). Applying the attractor field model to social cognition: Perceptual discrimination is facilitated, but memory is impaired for faces displaying evaluatively congruent expressions. *Journal of Personality and Social Psychology*, *93*(3), 335.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378.

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology*, *83*(5), 1076–1094.

Dai, H., & Micheyl, C. (2009). Psychophysical reverse correlation with multiple response alternatives. *Journal of Experimental Psychology-Human Perception and Performance*.

Darwin, C. (1887/1950). *Charles Darwin's autobiography*. Schuman.

Dawkins, R. (1976). *The selfish gene*. Oxford, UK: Oxford University Press.

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology*, *46*(5), 991–1004.

Der-Karabetian, A., & Smith, A. J. (1977). Sex-role stereotyping in the United States: Is it changing? *Sex Roles*, *3*(2), 193–198.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, *56*(1), 5–18.

Dijksterhuis, A. (2010). Automaticity and the unconscious. In S. T. Fiske,

D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 2, p. 228-267). Hoboken, NJ: John Wiley & Sons.

Dijksterhuis, A., & van Knippenberg, A. (1999). On the parameters of associative strength: Central tendency and variability as determinants of stereotype accessibility. *Personality and Social Psychology Bulletin*, *25*(4), 529.

Dotsch, R., Todorov, A., & Wigboldus, D. H. J. (in prep). A reverse correlation replication of primary dimensions of face evaluation.

Dotsch, R., & Wigboldus, D. H. J. (2008). Virtual prejudice. *Journal of Experimental Social Psychology*, *44*(4), 1194–1198.

Dotsch, R., Wigboldus, D. H. J., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978–980.

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (under review). Visual stereotype formation: Behavioral information biases facial representations of new groups.

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (under revision). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*.

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*(1), 22–37.

Dovidio, J. F., & Gaertner, S. L. (1986). Prejudice, discrimination, and racism: Historical trends and contemporary approaches. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (p. 1-34). Orlando, FL: Academic Press.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in experimental social psychology*, *36*, 1–52.

Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 2,

p. 1084-1121). Hoboken, NJ: John Wiley & Sons.

Dovidio, J. F., Kawakami, K., & Gaertner, S. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68.

Dovidio, J. F., Kawakami, K., Smoak, N., & Gaertner, S. L. (2008). The Nature of Contemporary Racial Prejudice. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 165–192). New York: Psychology Press.

Dubow, S. (1992). Fascism and Afrikanerdom. *Journal of African History*, *33*(3), 513–515.

Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European review of social psychology*, *5*(1), 1–35.

Ebner, N. C. (2008). Age of face matters: Age-group differences in ratings of young and old faces. *Behavior Research Methods*, *40*(1), 130.

Ekman, P. (2003). *Emotions revealed*. Times Books.

Feldman, J. (1988). Objects in categories and objects as categories. In T. K. Srull & R. S. Wyer Jr. (Eds.), *Advances in social cognition: A dual process model of impression formation* (pp. 53–63). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357–411). Boston: McGraw-Hill.

Fiske, S. T. (2005). Social cognition and the normality of prejudgment. In J. F. Dovidio & P. G. . L. A. Rudman (Eds.), *On the nature of prejudice: Fifty years after allport* (pp. 36–53). Malden, MA: Blackwell Publishing.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902.

Fiske, S. T., & Linville, P. W. (1980). What does the schema concept buy us? *Personality and Social Psychology Bulletin*, *6*(4), 543–557.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in experimental social psychology*, *23*, 1–74.

Fiske, S. T., Neuberg, S. L., Beattie, A. E., & Milberg, S. J. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, *23*(5), 399–427.

Fiske, S. T., & Pavelchak, M. A. (1986). Category-based versus piecemeal-based affective responses: Developments in schema-triggered affect. In R. M. Sorrentio & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 1, pp. 167–203). New York: Guilford.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. New York: Random House.

Fletcher-Watson, S., Findlay, J., Leekam, S., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*, 571–583.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press, USA.

Ford, T. E., & Stangor, C. (1992). The role of diagnosticity in stereotype formation: Perceiving group means and variances. *Journal of Personality and Social Psychology*, *63*(3), 356–367.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692.

Gibson, J. (1979). The ecological approach to perception. *Boston: Haughton Mifflin*.

Gilbert, D., & Hixon, J. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517.

Gluszek, A., & Dovidio, J. F. (2010). The Way They Speak: A Social Psychological Perspective on the Stigma of Nonnative Accents in Communication. *Personality and Social Psychology Review*, *14*(2), 214.

Gordijn, E. H., Koomen, W., & Stapel, D. A. (2001). Level of prejudice in relation to knowledge of cultural stereotypes. *Journal of Experimental Social Psychology*, *37*(2), 150–157.

Gosselin, F., Bacon, B. A., & Mamassian, P. (2004). Internal surface representations approximated by reverse correlation. *Vision research*, *44*(21), 2515–2520.

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, *14*(5), 505.

Greenwald, A. G. (1990). What cognitive representations underlie social attitudes. *Bulletin of the Psychonomic Society*, *28*(3), 254–260.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Greenwald, A. G., Pickrell, J. E., & Farnham, S. D. (2002). Implicit partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology*, *83*(2), 367–379.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments*. *Journal of Experimental Social Psychology*, *12*(4), 392–407.

Hamilton, D. L., & Sherman, S. J. (1989). Illusory correlations: Implications for stereotype theory and research. *Stereotyping and prejudice: Changing conceptions*, 59–82.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psy-*

chological Review, *103*(2), 336–355.

Hamilton, D. L., & Trolier, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. *Prejudice, discrimination, and racism*, 127–163.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*(5), 837–852.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223–232.

Higgins, E. T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 75–123). New York: Guilford.

Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. *Annual review of psychology*, *38*(1), 369–425.

Hilton, J., & Von Hippel, W. (1996). Stereotypes. *Annual review of psychology*, *47*(1), 237–271.

Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition*, *54*(2), 209–252.

Huart, J., Corneille, O., & Becquart, E. (2005). Face-based categorization, context-based categorization, and distortions in the recollection of gender ambiguous faces. *Journal of Experimental Social Psychology*, *41*(6), 598–608.

Hugenberg, K. (2005). Social categorization and the perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion*, *5*(3), 267–276.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*(6), 640–643.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization. *Psychological Science*, *15*(5), 342.

Hugenberg, K., & Sacco, D. F. (2008). Social categorization and stereotyping:

How social categorization biases person perception and face memory. *Social and Personality Psychology Compass*, *2*(2), 1052–1072.

Hutchings, P. B., & Haddock, G. (2008). Look Black in anger: The role of implicit prejudice in the categorization and perceived emotional intensity of racially ambiguous faces. *Journal of Experimental Social Psychology*, *44*(5), 1418–1420.

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (in prep). Visualizing ingroup projection - revealing "typical" european looks in a reverse correlation design.

Ishai, A. (2008). Let's face it: it's a cortical network. *NeuroImage*, *40*(2), 415–419.

Johnston, L., & Macrae, C. (1994). Changing social stereotypes: The case of the information seeker. *European Journal of Social Psychology*, *24*(5), 581–592.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., et al. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, *9*(4), 462–475.

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review*, *98*(1), 54–73.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*, 759–763.

Karremans, J., Dotsch, R., & Corneille, O. (in prep). Beauty blindness: Romantic involvement reduces sensitivity to opposite-sex attractiveness.

Klatzky, R. L. (1984). Visual memory: Definitions and functions. In R. S. Wyer, T. K. Srull, & J. Hartwick (Eds.), *Handbook of social cognition* (Vol. 2, pp. 233–270). Lawrence Erlbaum Associates.

Klatzky, R. L., Martin, G. L., & Kane, R. A. (1982). Influence of social-category activation on processing of visual information. *Social Cognition*, *1*(2), 95–109.

Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account of the Implicit Association Test effect. *Personality and Social Psy-*

chology Bulletin, *31*(2), 208.

Krueger, J. (1996). Personal beliefs and cultural stereotypes about racial characteristics. *Journal of Personality and Social Psychology*, *71*(3), 536–548.

Langner, O., Becker, E. S., & Rinck, M. (2009). Social Anxiety and Anger Identification. *Psychological Science*, *20*(6), 666.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D., Hawk, S., & Knippenberg, A. van. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*.

Lavater, J. C. (1772/1850). *Essays on Physiognomy: designed to promote the knowledge and the love of mankind*. W. Tegg.

Le Pelley, M. E., Reimers, S. J., Calvini, G., Spears, R., Beesley, T., & Murphy, R. A. (2010). Stereotype Formation: Biased by Association. *Journal of Experimental Psychology: General*, *139*(1), 138–161.

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, *135*(4), 501.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics*, *62*(2), 386–401.

Lewicki, P. (1986). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 135–146.

Lingle, J. H., Altom, M. W., & Medin, D. L. (1984). Of Cabbages and Kings: Assessing the Extendibility of Natural Object Concept Models to Social Things. In R. S. Wyer, T. K. Srull, & J. Hartwick (Eds.), *Handbook of social cognition* (p. 71-117). Lawrence Erlbaum Associates.

Lippmann, W. (1922). *Public opinion.* Nueva York, Free Press.

Liu, C. H., & Chaudhuri, A. (2003). What determines whether faces are special? *Visual Cognition*, *10*(4), 385–408.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-

based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*(1), 5–18.

Lombroso, C. (1876/2006). *Criminal man*. Duke Univ Pr.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental scores. *Reading, Addison–Wesley*.

Lundqvist, D., & Litton, J. E. (1998). *The Averaged Karolinska Directed Emotional Faces-AKDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet* (Tech. Rep.). ISBN 91-630-7164-9.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual review of psychology*, *51*, 93–120.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology*, *69*(3), 397–407.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Calvini, G. (1999). Seeing More Than We Can Know: Visual Attention and Category Activation* 1. *Journal of Experimental Social Psychology*, *35*(6), 590–602.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. J., & Castelli, L. (1997). On the activation of social stereotypes: The moderating role of processing objectives. *Journal of Experimental Social Psychology*, *33*, 471–489.

Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, *23*(1), 77–87.

Macrae, C. N., & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology*, *37*(5), 793–805.

Macrae, C. N., & Quadflieg, S. (2010). Perceiving people. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, p. 428-463). Hoboken, NJ: John Wiley & Sons.

Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality*

and Social Psychology Review, *8*(4), 383.

Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of Black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin*, *28*(2), 250.

Madon, S. (1997). What do people believe about gay males? A study of stereotype content and strength. *Sex Roles*, *37*(9), 663–685.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, *28*(2), 209–226.

Martin-Malivel, J., Mangini, M. C., Fagot, J., & Biederman, I. (2006). Do humans and baboons use the same information when categorizing human and baboon faces? *Psychological Science*, *17*(7), 599.

Mason, M. F., Cloutier, J., & Macrae, C. N. (2006). On construing others: Category and stereotype activation from facial cues. *Social Cognition*, *24*(5), 540–562.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215–238.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442.

McConnell, A. R., Rydell, R. J., Strain, L. M., & Mackie, D. M. (2008). Forming implicit and explicit attitudes toward individuals: Social group association cues. *Journal of Personality and Social Psychology*, *94*(5), 792.

McGarty, C. (1999). *Categorization in social psychology*. Thousand Oaks, CA: SAGE Publications.

McGarty, C. (2002). Stereotype formation as category formation. In C. McGarty, V. Y. Yzerbyt, & R. Spears (Eds.), *Stereotypes as explanations: The formation of meaningful beliefs about social groups* (p. 16-37). University Press, Cambridge.

McGarty, C., Yzerbyt, V. Y., & Spears, R. (2002). *Stereotypes as explanations: The*

formation of meaningful beliefs about social groups (C. McGarty, V. Y. Yzerbyt, & R. Spears, Eds.). University Press, Cambridge.

McNeill, D. (1998). *The face*. Little, Brown, Boston.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*(1), 113–138.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*(1), 89–115.

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*(6), 1180–1192.

Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology*, *29*(1), 11–27.

Mummendey, A., & Wenzel, M. (1999). Social discrimination and tolerance in intergroup relations: Reactions to intergroup difference. *Personality and Social Psychology Review*, *3*(2), 158.

Neisser, U. (1987a). *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge University Press Cambridge, UK.

Neisser, U. (1987b). From direct perception to conceptual structure. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 11–24). Cambridge University Press Cambridge, UK.

Nestor, A., & Tarr, M. J. (2008). Gender recognition of human faces using color. *Psychological Science*, *19*(12), 1242.

Oakes, P. J. (1987). The salience of social categories. In P. J. O. S. D. R. . M. S. W. J. C. Turner M. A. Hogg (Ed.), *Rediscovering the social group: A self-categorization theory* (p. 117-141). Oxford: Basil

Blackwell.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872.

Ostrom, T. (1984). The sovereignty of social cognition. In R. S. Wyer, T. K. Srull, & J. Hartwick (Eds.), *Handbook of social cognition* (Vol. 1, pp. 1–38). Lawrence Erlbaum Associates.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987.

Penton-Voak, I. S., Jacobson, A., & Trivers, R. (2004). Populational differences in attractiveness judgements of male and female faces:: Comparing British and Jamaican samples. *Evolution and Human Behavior*, *25*(6), 355–370.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, *394*(6696), 884–887.

Pettigrew, T. F., Allport, G. W., & Barnett, E. O. (1958). Binocular resolution and perception of race in South Africa. *British Journal of Psychology*, *49*(4), 265–278.

Pinter, B., & Greenwald, A. G. (2005). Clarifying the Role of the "Other" Category in the Self-Esteem IAT. *Experimental Psychology*, *52*(1), 74–79.

Quinn, K. A., & Macrae, C. N. (2005). Categorizing others: The dynamics of person construal. *Journal of Personality and Social Psychology*, *88*(3), 467–479.

Ratliff, K. A., & Nosek, B. A. (2010). Creating distinct implicit and explicit attitudes with an illusory correlation paradigm. *Journal of Experimental Social Psychology*, *46*(5), 721–728.

Rees, G., Frith, C., & Lavie, N. (2001). Processing of irrelevant visual motion

during performance of an auditory attention task. *Neuropsychologia*, *39*(9), 937–949.

Rennels, J. L., Bronstad, P. M., & Langlois, J. H. (2008). Are attractive men's faces masculine or feminine? The importance of type of facial stimuli. *Journal of Experimental Psychology-Human Perception and Performance*, *34*(4), 884–892.

Rhodes, G. (2006). The evolution of facial attractiveness. *Annual Review of Psychology*, *57*, 199–226.

Rhodes, G., Hickford, C., & Jeffery, L. (2000). Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? *British Journal of Psychology*, *91*(1), 125–140.

Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, *28*(2), 147–166.

Rojahn, K., & Pettigrew, T. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology*, *31*(2), 81–109.

Rosch, E., & Lloyd, B. (1978). *Cognition and categorization*. Lawrence Erlbaum Associates.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, *7*(4), 573–605.

Rossion, B. (2008). Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. *NeuroImage*, *40*(2), 423–426.

Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145–181). Lawrence Erlbaum Associates.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*(2), 79.

Rudman, L. A., Phelan, J. E., & Heppen, J. B. (2007). Developmental sources of implicit attitudes. *Personality and Social Psychology Bulletin*, *33*(12), 1700.

Rule, N. O., & Ambady, N. (2008a). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*,

44(4), 1100–1105.

Rule, N. O., & Ambady, N. (2008b). The face of success. *Psychological Science*, *19*(2), 109.

Rule, N. O., Ambady, N., Adams Jr, R. B., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of personality and social psychology*, *95*(5), 1019–1028.

Ruys, K. I., Dijksterhuis, A., & Corneille, O. (2008). On the (mis) categorization of unattractive brides and attractive prostitutes: Extending evaluative congruency effects to social category activation. *Experimental psychology*, *55*(3), 182–188.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995.

Said, C., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and non-social face dimensions. *Neuropsychologia*.

Scott, L. S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science*, *20*(6), 676.

Secord, P. F., Bevan, W., & Katz, B. (1956). The Negro stereotype and perceptual accentuation. *Journal of Abnormal and Social Psychology*, *53*(1), 78–83.

Secord, P. F., Dukes, W. F., & Bevan, W. (1954). Personalities in faces: I. An experiment in social perceiving. *Genetic Psychology Monographs*, *49*, 231–270.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology*, *70*, 1126–1141.

Sherman, J. W., & Hamilton, D. L. (1994). On the formation of interitem associative links in person memory. *Journal of Experimental Social Psychology*, *30*(3), 203–217.

Silvern, L. E., & Ryan, V. L. (1983). A reexamination of masculine and feminine sex-role ideals and conflicts among ideals for the man, woman, and person. *Sex Roles*, *9*(12), 1223–1248.

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology*, *77*(5), 885–904.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press Cambridge, MA.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108.

Smith, E. R., & Zarate, M. A. (1990). Exemplar and prototype use in social categorization. *Social Cognition*, *8*(3), 243–262.

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*(1), 3–21.

Smith, M. L., Cottrell, G. W., Gosselin, F. A., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*(3), 184.

Solomon, J. A. (2002). Noise reveals visual mechanisms of detection and discrimination. *Journal of Vision*, *2*(1).

Stangor, C., & Duan, C. (1991). Effects of multiple task demands upon memory for information about social groups. *Journal of Experimental Social Psychology*, *27*(4), 357–378.

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*(2), 207–218.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*(1), 42–61.

Stangor, C., & Schaller, M. (1996). *Foundations of stereotypes and stereotyping*. Guilford New York.

Steffens, M. C., & Buchner, A. (2003). Implicit Association Test: Separating

transsituationally stable and variable components of attitudes toward gay men. *Experimental Psychology*, *50*(1), 33–48.

Stroebe, W., & Insko, C. A. (1989). Stereotypes, prejudice, and discrimination: Changing conceptions in theory and research. In D. Bar-Tal, C. F. Grauman, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotypes and prejudice: Changing conceptions* (pp. 3–34). New York: Springer.

Tajfel, H. (1969). Cognitive Aspects of Prejudice. *Journal of social Issues*, *25*(4), 79–97.

Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*, 764–770.

Taylor, S. E., & Crocker, J. (1981). Social cognition. In E. T. Higgins, P. C. Herman, & M. P. Zanna (Eds.), *Schematic bases of social information processing* (p. 89-134). Hillsdale, NJ: Erlbaum Associates.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*(7), 778–793.

Todorov, A. (2008). Evaluating Faces on Trustworthiness. *Annals of the New York Academy of Sciences*, *1124*(The Year in Cognitive Neuroscience 2008), 208–224.

Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (under review). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass*.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in cognitive sciences*, *12*(12), 455–460.

Turner, J. C. (1987). A self-categorization theory. In J. C. Turner, M. A. Hogg,

P. J. Oakes, S. D. Reicher, & M. S. Wetherell (Eds.), *Rediscovering the social group: A self-categorization theory* (p. 42-67). Oxford: Basil Blackwell.

van Knippenberg, A., & Dijksterhuis, A. (2000). Social categorization and stereotyping: A functional perspective. *European Review of Social Psychology*, *11*(1), 105–144.

van Knippenberg, A., van Twuyver, M., & Pepels, J. (1994). Factors affecting social categorization processes in memory. *British Journal of Social Psychology*, *33*, 419–419.

van Twuyver, M., & van Knippenberg, A. (1995). Social categorization as a function of priming. *European journal of social psychology*, *25*(6), 695–701.

Verkuyten, M., & Zaremba, K. (2005). Interethnic relations in a changing political context. *Social Psychology Quarterly*, *68*(4), 375.

Victor, J. D. (2005). Analyzing receptive fields, classification images and functional images: challenges with opportunities for synergy. *Nature neuroscience*, *8*(12), 1651–1656.

Vonk, R., & van Knippenberg, A. (1995). Processing attitude statements from in-group and out-group members: Effects of within-group and within-person inconsistencies on reading times. *Journal of Personality and Social Psychology*, *68*(2), 215–227.

Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology*, *27*(4), 635–657.

Wigboldus, D. H. J., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology*, *78*(1), 5–18.

Wiggett, A. J., & Downing, P. E. (2008). The Face Network: Overextended?(Comment on. *NeuroImage*, *40*(2), 420–422.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598.

Wilson, D., & Stevenson, R. (2003). The fundamental role of memory in olfactory perception. *TRENDS in Neurosciences*, *26*(5), 243–247.

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science*, *17*(9), 799.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262–274.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, *15*(8), 1234.

Zebrowitz, L. A. (1996). Physical appearance as a basis of stereotyping. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 79–120). The Guilford Press.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Westview Pr.

Zebrowitz, L. A. (2006). Finally, faces find favor. *Social Cognition*, *24*(5), 657–701.

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology*, *28*, 1143–1143.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517.

Nederlandse samenvatting

Het waarnemen van andere mensen in het dagelijks leven is een complex proces. Het lijkt zo simpel: je ziet iemand over straat lopen en je weet zonder na te denken of de persoon man of vrouw is, jong of oud, Marokkaans of Nederlands (Brewer, 1988). Dit proces, waarin personen die je waarneemt automatisch in relevante categorieën worden ingedeeld, heet sociale categorisatie (Allport, 1954). Mensen worden voornamelijk in termen van hun categorielidmaatschap waargenomen in plaats van hun individuerende kenmerken. Dat heeft voordelen, omdat gaten in kennis over de waargenomen persoon opgevuld worden met kennis over de categorie (het stereotype). Zo kunnen mensen makkelijk op basis van slechts weinig informatie besluiten hoe ze omgaan met de waargenomen persoon: kom ik dichterbij of vermijd ik deze persoon? Sociaal psychologen hebben in de afgelopen zestig jaar veel onderzoek verricht naar dit proces (Tajfel, 1969), dat de basis vormt van vooroordelen en discriminatie.

Hoewel veel bekend is over de gevolgen van sociale categorisatie, is er bij sociaal psychologen vrij weinig bekend over wat plaats vindt vóór sociale categorisatie. Het antwoord op de vraag hoe lichtgolven die op het netvlies vallen uiteindelijk tot de waarneming van bijvoorbeeld een jonge Marokkaanse

jongen leiden is nog onduidelijk. Op een bepaald punt in het waarnemingsproces in de hersenen dient de binnenkomende visuele informatie omgezet te worden in de kennis dat wat waargenomen wordt een persoon is die tot een bepaalde categorie behoort. Uit eerder onderzoek is al duidelijk welke factoren bepalen welke categorie geselecteerd wordt als meerdere categorieën van toepassing zijn (zie bijvoorbeeld Macrae & Quadflieg, 2010). Maar hoe bepaalt het cognitieve systeem welke categorieën überhaupt van toepassing zijn op de binnenkomende visuele informatie? Deze vraag is de kern van dit proefschrift.

Wanneer is een bepaalde categorie van toepassing op een waargenomen persoon? In de afgelopen decennia is gesuggereerd dat categorietoepasselijkheid afhangt van normatieve *fit* tussen een waargenomen persoon en een categorie (Bruner, 1957; Oakes, 1987; van Knippenberg & Dijksterhuis, 2000). Normatieve fit refereert aan de mate waarin eigenschappen van een persoon overeenkomen met de eigenschappen die een waarnemer van categorieleden verwacht. Deze verwachtingen zijn onderdeel van het stereotype (Stangor & Schaller, 1996). Hieruit volgt dat stereotypen betrokken zijn bij het vaststellen van normatieve fit en dus bij het indelen van personen in categorieën. Dat stereotypen deze rol spelen is nooit eerder geëxpliciteerd door sociaal psychologen, hoewel sommigen wel dichtbij kwamen (Blanz, 1999; van Knippenberg & Dijksterhuis, 2000; McGarty, 1999). Voor zover wij weten bestaat tot op heden geen empirisch werk over de invloed van stereotype-inhoud op de allocatie van sociale categorieën.

In dit proefschrift hebben we getracht de rol van stereotype-inhoud op de allocatie van sociale categorieën te beschrijven. We hebben deze rol getoetst aan de hand van het categoriseren van gezichten. Gezichten zijn een belangrijke bron van sociale informatie (Macrae & Quadflieg, 2010) en trekken onmiddellijk de aandacht wanneer een ander waargenomen wordt (Fletcher-Watson et al., 2008). In het door ons voorgestelde *uitgebreide* model van sociale categori-

satie stellen we dat stereotypen visuele informatie bevatten die het uiterlijk van typische groepsleden omschrijft (zie Figuur 1.2 op pagina 13). Dit voorstel is eigenlijk al heel oud; de oorspronkelijke definitie van het woord stereotype van Lippmann in 1922 was "pictures in our heads" (plaatjes in onze hoofden). In ons model wordt die visuele informatie door het categorisatieproces gebruikt als vergelijkingskader om normatieve fit vast te stellen. Hoe beter een waargenomen gezicht bij de visuele informatie in het stereotype behorend bij een bepaalde categorie past, des te beter de normatieve fit met die specifieke categorie, en des te groter de kans dat dat gezicht in die categorie ingedeeld wordt. Bijvoorbeeld, in Nederland heerst het stereotype over Marokkaanse jongeren dat ze crimineel zijn (Gordijn et al., 2001). Wij veronderstellen dat als gevolg hiervan het visuele stereotype over Marokkanen crimineel-uitziende gezichten bevat en dat daarom mensen verwachten dat Marokkaanse gezichten er crimineel uitzien en dus geneigd zijn crimineel-uitziende gezichten als Marokkaans te categoriseren.

Om het visuele stereotype over Marokkanen bloot te leggen hebben we een taak ontwikkeld op basis van een in de cognitieve psychologie reeds bestaande taak, een zogenaamde *psychophysical reverse correlation image classification* taak (Dotsch et al., 2008; Mangini & Biederman, 2004; Todorov et al., under review). In deze taak kregen proefpersonen herhaaldelijk twee met ruis bedekte gezichten te zien (zie Figuur 2.1 op pagina 24), waarvan zij aangaven welke van de twee het meest Marokkaans op hen over kwam. Feitelijk werd steeds hetzelfde gezicht getoond, met steeds andere ruis. De ruis vertekende de gezichten, waardoor het lijkt alsof elk gepresenteerd gezicht een andere is. Op basis van de keuzes die proefpersonen maken kan een computeralgoritme uitrekenen en visualiseren welke ruis op een gezicht geprojecteerd moet worden om aan de verwachtingen die een proefpersoon heeft van typisch Marokkaanse gezichten te voldoen. De uitkomsten hiervan zijn een indicatie van het visuele stereotype van proefpersonen (zie Figuur 2.2 op pagina 25 voor de visualisaties).

Niet iedereen heeft hetzelfde stereotype over Marokkanen. Terwijl sommigen Marokkanen sterk met criminaliteit associëren, zullen anderen die associatie veel minder sterk hebben. Met een impliciete associatie taak (IAT; Greenwald et al., 1998) is het mogelijk om op basis van reactietijden te schatten hoe sterk de negatieve associaties met Marokkanen bij een proefpersoon zijn (ook wel impliciete bevooroordeeldheid genoemd). Bevooroordeeldheid wordt door sociaal psychologen gedefinieerd als de mate waarin mensen negatief denken over een groep (Wittenbrink et al., 1997; Dovidio & Gaertner, 2010) en de IAT meet dat op indirecte wijze. De mate waarin proefpersonen bevooroordeeld zijn beïnvloedt de inhoud van stereotypen (Gordijn et al., 2001; Wittenbrink et al., 1997). Met andere woorden, de negatieve associaties met Marokkanen voorspellen in welke mate het stereotype van proefpersonen over Marokkanen negatieve kenmerken bevat. In de studies in Hoofdstuk 2 van dit proefschrift wordt aangetoond dat dit ook tot uiting komt in het visuele stereotype: proefpersonen die sterke negatieve associaties hadden met Marokkanen verwachtten dat typisch Marokkaanse gezichten er crimineler uitzagen.

Het lijkt dus zo te zijn dat mensen visuele stereotypen hebben over groepen en dat die visuele stereotypen per persoon kunnen verschillen. Het is de vraag waar deze visuele stereotypen vandaan komen. In Hoofdstuk 3 onderzochten we of visuele stereotypen kunnen ontstaan op basis van beschrijving van gedragingen van categorieleden. Tegelijkertijd konden we hier de voorgestelde causale relatie tussen impliciete bevooroordeeldheid en visueel stereotype aantonen. Deze relatie werd in Hoofdstuk 2 met bestaande groepen correlationeel aangetoond, maar onduidelijk bleef of impliciete vooroordelen visuele stereotypen beïnvloeden of andersom. Om de richting van deze relatie aan te tonen maakten we in Hoofdstuk 3 gebruik van stereotypen over een nieuwe groep die tot proefpersonen in het lab aankwamen niet bestond: Groep X. We gebruikten een stereotypeformatie paradigma gebaseerd op werk van Crawford et al. (2002), waarin aan proefpersonen gevraagd werd om een indruk te vor-

men van groep X op basis van gezichten en gedragsbeschrijvingen. De inhoud van de gedragsbeschrijvingen was gemanipuleerd. Voor de helft van de proefpersonen waren de beschreven gedragingen voornamelijk crimineel van aard, voor de andere helft waren deze voornamelijk betrouwbaar van aard. Beide groepen proefpersonen kregen wel allemaal dezelfde gezichten te zien. Achteraf werden zowel de visuele stereotypen over groep X gemeten (met een zelfde taak als in Hoofdstuk 2, met visualisaties van typische groep X gezichten als uitkomst) als de mate van bevooroordeeldheid.

Zoals verwacht waren proefpersonen in de criminele gedragsbeschrijvingen conditie meer bevooroordeeld over groep X (impliciet en expliciet) dan proefpersonen in de betrouwbare gedragsbeschrijvingen conditie. Bovendien zagen de gevisualiseerde visuele stereotypen van proefpersonen in de criminele conditie er crimineler uit dan die van proefpersonen in de betrouwbare conditie. De mate waarin de visuele stereotypen beïnvloed werden door de gedragsbeschrijvingen werd gemedieerd door bevooroordeeldheid. Deze bevindingen zijn om twee redenen belangrijk: 1) ze leveren het eerste experimentele (en daarmee causale) bewijs voor het effect van vooroordelen op de inhoud van het visuele stereotype en 2) ze suggereren dat de vertekeningen (*bias*) in visuele stereotypen niet het gevolg hoeven zijn van de echte gezichten van groepsleden die mensen zijn tegengekomen, maar juist het gevolg kunnen zijn van groepsinformatie ongerelateerd aan gezichten.

Nu vastgesteld is dat vooroordelen de inhoud van visuele stereotypen beïnvloeden, werden in Hoofdstuk 4 de implicaties voor sociale categorisatie onder de loep genomen. Ons model veronderstelt dat bevooroordeeldheid de fit tussen toegankelijke categorieën en binnenkomende visuele informatie beïnvloedt via de inhoud van visuele stereotypen. Gezichten die een betere fit hebben met het visuele stereotype van een bepaalde categorie hebben een grotere kans om als zodanig gecategoriseerd te worden. De studies in Hoofdstuk

4 laten zien dat dit inderdaad het geval is. Bovendien belichten de uitgevoerde studies de precieze relatie tussen bevooroordeeldheid en het visuele stereotype: bevooroordeelheid voorspelt de mate waarin het visuele stereotype gezichtskenmerken bevat die met negatieve of positieve persoonlijkheidstrekken geassocieerd worden. Daarnaast kan het visuele stereotype gezichtskenmerken bevatten die met neutrale persoonlijkheidstrekken geassocieerd worden, maar dat wordt niet voorspeld door bevooroordeeldheid. Welke specifieke persoonlijkheidstrekken gepresenteerd worden in het visuele stereotype is niet noodzakelijk gerelateerd aan bevooroordeeldheid, maar kan bijvoorbeeld afhangen van welke trekken een voorspellende waarde hebben voor groepslidmaatschap (zie Le Pelley et al., 2010).

De studies in Hoofdstuk 4 bieden bewijs voor deze relatie. De persoonlijkheidstrek *crimineel* is stereotypisch voor Marokkanen en is daarom onderdeel van het visuele stereotype (zoals in Hoofdstuk 2 was aangetoond). Omdat crimineel een negatieve trek is voorspelt bevooroordeeldheid de mate waarin de trek onderdeel is van het visuele stereotype en uiteindelijk de kans dat een crimineel-uitziend gezicht als Marokkaans wordt gecategoriseerd (Studie 4.1 en 4.2). Andere negatieve trekken, zoals *dom*, zijn niet stereotypisch voor Marokkanen (Gordijn et al., 2001) en zijn daarom geen deel van het visuele stereotype. Een gevolg hiervan is dat bevooroordeeldheid niet gerelateerd is aan de kans dat gezichten die er dom uitzien als Marokkaans gecategoriseerd worden (Studie 4.2). Dit is belangrijk, omdat dit aantoont dat iemand die bevooroordeeld is (negatief denkt over een groep) niet blind alle negatieve gezichten in die groep indeelt. Alleen wanneer de negativiteit gebaseerd is op gezichtskenmerken die geassocieerd zijn met stereotypische persoonlijkheidskenmerken, beïnvloedt bevooroordeeldheid de kans dat een gezicht bij de negatieve groep wordt ingedeeld.

Tot slot laten we in Hoofdstuk 4 zien dat ons model ook opgaat voor andere sociale categorieën. In Studie 4.3 toetsen we ons model met behulp van de categorie homoseksueel. "Vrouwelijk" is een stereotypische eigenschap voor homoseksuelen en daarom zijn vrouwelijke gezichtskenmerken deel van het visuele stereotype. Omdat het kenmerk positief is, is de kans juist kleiner dat vrouwelijke gezichten als homoseksueel gecategoriseerd worden naar mate proefpersonen sterker negatief bevooroordeeld zijn. Ter ondersteuning van de hypothese dat alleen valentie-geladen kenmerken die geassocieerd zijn met stereotypische eigenschappen deel uitmaken van visuele stereotypen, laten we in Studie 4.3 bovendien zien dat negatief bevooroordeelde proefpersonen niet meer de neiging hebben om crimineel-uitziende gezichten als homoseksueel te categoriseren dan minder sterk bevooroordeelde proefpersonen.

Sociaal psychologen hebben zich tot op heden voornamelijk beziggehouden met de *gevolgen* van sociale categorisatie. Zodra een waargenomen persoon gecategoriseerd is, wordt aan de categorie gerelateerde informatie actief: het stereotype. Het model dat we in dit proefschrift hebben geïntroduceerd veronderstelt echter nog een tweede functie van het stereotype. Het stereotype beïnvloedt namelijk of een waargenomen persoon wel of niet bij een categorie past. Het empirisch werk in dit proefschrift ondersteunt dit model indirect, door een effect van bevooroordeeldheid op visuele stereotype-inhoud en een effect van bevooroordeeldheid op sociale categorisatie aan te tonen. Omdat het specifieke patroon van het laatste effect pas begrijpelijk wordt in het licht van visuele stereotypen lijkt de tweede functie van het stereotype aannemelijk. Dit proefschrift is hiermee het eerste werk dat aantoont dat het sociale categorisatie proces, waarvan in de loop der jaren zoveel vertekenende gevolgen zijn gedocumenteerd, zelf slachtoffer is van vertekeningen.

Acknowledgements

> "A dream you dream alone is only a dream. A dream you dream together is reality." - John Lennon

This dissertation is not the product of one person alone. I have many people to be grateful to.

Daniël Wigboldus, you sparked my interest in scientific research. I have fond memories of our collaboration, even from before I started as a PhD student: running our first virtual reality experiment in the CAVE in Amsterdam, setting up the RIVERlab in Nijmegen, our conversations in the train while commuting, and working on grant proposals in your home. You are my mentor in science and academia. Your insights have been invaluable to my work and career. Above all, I consider you to be a great friend, with whom I share many interests. I hope our paths will lead us in similar directions for many years to come.

Ad van Knippenberg, strategic mind, masterful debater, gatekeeper of science, godfather of social cognition, and polarization filter. I always knew that if I could convince you of an argumentation, it would hold during my defense.

However, I am still not sure I ever convinced you. You taught me how to contribute to science in meaningful ways while keeping close to data. You pointed out to me that social psychologists have actually conducted research before the last two decades. And, most importantly, you challenged me to dive below the surface.

Daniël and Ad, I could not have wished for better advisors. The fact that we had to close the door for our meetings, as to not disturb other people on the ninth floor, was purely consequence of intellectual disagreement between passionate researchers. I hope we keep working together despite future distances and retirements, although this time maybe without Ad's rule of thumb: "By default, I agree with Daniël, unless he is wrong".

Alexander Todorov, after we met in Nijmegen, we immediately were a good match in terms of research interests and research methods. After the warm welcome in your lab, I learned a lot from your way of working. I am very much looking forward to an inspiring and productive collaboration next year, in which I can fully focus on our joint research projects.

I am very grateful to those who (sometimes without being aware of it) have inspired or challenged me to ask bigger, deeper, and more daring questions: Ap Dijksterhuis, Jack Dovidio, Willem Frankenhuis, Rob Holland, Johan Karremans, and Rick van Baaren.

Martijn and Peggy, thank you for making my move to Nijmegen very easy. I will never forget your warm hospitality (and all the times I came to shower at your home when my own bathroom was being renovated). I am happy we have become good friends.

Although I liked working with all my colleagues on the ninth floor, there are several colleagues that I am particularly grateful to: Oliver, our conversations on statistics, programming, and computer games might come across as very nerdy, but I loved them. I learned a lot from you and I am glad you were my office mate. I hope you didn't mind my sarcastic humor too much. I am sure we will keep working together, despite the both of us moving farther away from Nijmegen; Mattheis, in trying to mimic the lifestyle of passionate scientists, we started to work during and after business hours in cafe's. I enjoyed our discussions about science, academia, and politics very much. However, I still don't like outdoor activities; Gijs, thank you for being a friend besides being a colleague. Your support helped me a lot; Tila, the only PhD student of my generation, we kick ass; Annemarie, Aukje, Lieke, Thijs, Rui, and Jeroen thank you for making work fun. I am going to miss being around you all.

I am grateful to Elsje and Michael Reiss, who made my visit to Princeton possible by offering me a bedroom to sleep in and their house to live in. For two months I have been a part of your family, thank you for taking me in. As promised, I will pay it forward.

Working on this dissertation has been a smooth ride because of Ronny, Marijke, Marjo, Thea, Saskia, Pascal, Hubert, and Giti. Thank you for your help.

When I moved from Amsterdam to Nijmegen to work on my dissertation, I left behind some very special people. Friends with whom I grew up, studied, played music and with whom I still have drinks, game, go on vacation, and sometimes even collaborate. Friends who got married, pregnant, or promoted and whom I was happy to take a break from my dissertation for. I am grateful that after four years I can still call all of you my friends, although I definitely do not see you often enough. I am confident that our friendship will stay strong when I am away for one additional year, but I will miss you all the more. In the end, I intend to move near you again.

There is one friend I want to mention in particular. Maarten, thank you for being the best friend someone could wish for. You are family to me. Thank you for challenging my conventional way of thought. We lead such different lives, but I believe us to be as similar as we are different. Now that we both have finished our dissertations, when are we going to take over the world?

Tom, you are both family and a great friend. Thank you for the conversations only brothers can have and your support in the last three decades. I am proud to call you my brother and I am honored that you are my paranymf.

Papa, mama, I know how proud you are of me. I am who I am today because of you. You always encouraged me to explore and develop. Of course it's easy to say from my perspective, but I think you have done a wonderful job raising Tom and me. Aba, ima, ve achi, ani ohev otchem be col ha lev.

Christine, my girl, my muse, you supported me all the way, unconditionally. Thank you for being there for me and taking care of me. And thank you for being part of my family, as I have become part of your very supportive and interested family. The one thing I regret most about moving to Nijmegen is that I did it without you. John Lennon was right when he said that "as usual, there is a great woman behind every idiot". I want to dream my next dream together with you in New York. I love you.

Nijmegen, November 16th, 2010

Curriculum vitae

Ron Dotsch was born on June 11th, 1980 in Amstelveen, the Netherlands. After finishing primary education at the Martin Luther King school in Amstelveen, he went on to the Vossius Gymnasium in Amsterdam. In 1998 he started studying Artificial Intelligence at the University of Amsterdam. After a short intermezzo to get his Black Jack dealer degree, he made the switch to Psychology at that same university. During his study he chaired the Student Association for Psychology Students for one year. He graduated cum laude in Social Psychology on the use of virtual reality as a social psychological research tool. He then built the Nijmegen virtual reality lab (RIVERlab) together with Daniël Wigboldus. In 2006 he started to work on his PhD project on face categorization and prejudice, supervised by Daniël Wigboldus and Ad van Knippenberg at Radboud University Nijmegen. In 2011 he will start working as a postdoc researcher at Princeton, New Jersey, supervised by Alexander Todorov.